


```

APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hofle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
FILE REFERENCE: PCT/US 99/23535
CURRENT FILING DATE: 1999-10-07
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: Patentn Ver. 2.1
SEQ ID NO 79
LENGTH: 1213
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-79

```

```

Alignment scores:
Quality: 119.50      Length: 491
Ratio: 0.660        Gaps: 22
Percent Similarity: 36.864      Percent Identity: 25.051

```

Alignment block:

US-09-303-518D-127 x US-09-413-814-79 ..

Align seg 1/1 to: US-09-413-814-79 from: 1 to: 1213

```

18 AGGTCTAAACCTGCCCATCGGGGACAGCGGACGAATGATTTATGACG 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 ArgArgAlaProAlaAlaArgArgLysHisGlyArgProGlyLeu...Ar 132
68 GGGCCGTCATTCGAGAGTCGCGTGGTGGCGGACGAATATGCGGGTAG 117
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 GasPLeuHisValArgValHisGlyAla..... 141
118 GCGCCCTNGATGAAGTCAAGGAAGCGGATCGGCAAAAAGGCCAAGT 167
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 .....AlaGlnGlyArgAspArgProProArg..... 151
168 GCTGTGAAGACAAAGNATCGGGCGGTGTGTACCGCGCGCTTT 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
152 .....ArgGlyLeuHisGlyProArgHis 159
218 CAGGCAAAAT...CGCGCCCATCATCGGCGGAAAGCGCTACTTCA 264
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
159 sGlnProProLeuArgArgArgProGlyGlyProGlyAlaArgAlaLeu 176
265 TCGGT.....CGTATTGCGCGTTGA 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 AlaAlaGluLeuArgProValGlyLeuArgValArgAspAlaArg 192
285 AGGCAAGCAAGAAATCGAGTTGCAACGCTACCGCGGAGGCTTGA 334
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
193 ArgArgArgArgArgAspArgProArgProArgGlyLeuGlySerGlyAl 209
335 ACTTAAGCGCGGANGAANTNNGCATCTGATCCCAATCCGGTTTGG 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209 AlaLeuAlaArgAlaArgGlyAlaArgAlaGlyAspArgValGlyLeu 226
385 ACTGCGCTCGTANCGCTCGTTCAAGCAAAATCCCTGCGGT...CGAT 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 roGlyAlaAspArgLysAspAlaHisGlyArgValProArgArgGlyArg 242
432 CGAGCCGTT.....CGCAT...CTTCGTCATCGCATGACACA 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
243 GlyAlaValAlaValArgProArgHisAspArgLysArgGlyLeuAsp 259
467 CCAATCCGCTNCGGACAGCCCTGTGTTGATCAAAAGAACCGGCGAN 516
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

259 acGluAlaProArgSerHisProArgGlyLeuProArgAlaProArgArg 276
517 GATTTCAGACGANGTTCGTGATTTAG.....CGTTTGAC 554
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 LeuProArgArgArgArgArgGlyAlaAspLeuValAspArgProArg 292
555 CAGCGCTAAATTCATGTGTGTAAAGGACGTGGCGGACGCGCTGTG 604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 ArgGlyArgArgProGlyVal..... 299
605 AAAATGTCGCAACATCGAACAATTCGCGCGCGCGGATCGGCGC 654
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 .....AlaGlnHisPro.....LeuArgProArg.....A 309
655 GGTTCGATGCGGACGACGATTCATTCATTCGAGCGGCTGCGTCAAA... 701
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 rGlyProAlaHisLeuArgAlaArgArgGlyAlaGlyAlaValSerAsp 325
701 ..... 701
326 ArgGlyProArgArgAspProHisArgArgAspArgArgAlaArg 342
701 ..... 701
342 eLeuAlaArgArgGlyAlaAspProGlyAlaValProGlyAlaProHis 359
702 .....CAAAACGTTTGACCATCATTCATTCAGATGT 734
359 sPArgArgAlaAlaValGlnAspArgArgProGlyProLeuLeuArg 375
735 AATTCGATCGGAGCTTTGTTGCAACAGCCCTCT..... 770
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
376 ArgAspHisArgAlaAlaArgAlaHisArgProSerGlyAlaAspArg 392
771 .....GAACACGACGCGGTGATTCCTTGGGTGTTCTCAAGTCAACA 816
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 gLeuProHisArgAlaArgArgAspArgGly.....ArgProArg 406
817 CCAAGGCTCTTGCGTACCGCTTTGGGTGCGAAAGATGCGAATTAAT 866
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
406 AlaAlaProLeu..... 409
867 GGGCGAATTCGTTGACGCGACACCGCGTGATTCGGTTCGATTTGA 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 GlyArgAlaGlyGlyArg.....GlyGlyGly 418
917 ACGGCGGATTCACAAAGCGCGGACGATTTATTCGACGCTACCAAT 966
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 uAspArgSerValGlyArgGlyAla.....ProGly 429
967 CAGATTCGTTATCGAAGAGCGCGCAAGAGAGCTGTCGGCTGGGT 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 rGlyAlaArgArgArgArgArgArgArgArgArgArgArgProLeu... 444
1017 TGGCGCGGACGCGCAAAATCTCATGACGCTGACGCGCTGCGCAT 1066
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445 .....ArgPro... 446
1067 TCCTGAACAAACCTTTCAAGTTTCAGACAGACCGCTCAAGCGTGG 1116
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
447 .....GluGluAlaAlaGlyValHisAspProArgGlyGlyArgArg 461
1117 GCGCGCATGTCGCGATTCGATTCAGAGCGCGTAAAGCGCGTACAT 1166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
461 ro...ProGlyAlaAlaAlaGluArgGluAlaArgGlyGlyProArg 476
1167 CCGTACCTACCTGCTTTGCGCGATTTAATGTCGG..... 1202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
477 ProAlaGlySerArgArgGlyArgAlaAlaArgArgGlyArgArgPro 493
1203 .....CGATACGACAGCGCGGACGATGTTGGTGTG 1236
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
493 aAspArgAspArgAlaAlaHisArgLeuGlyAlaArgGlyLeuAlaG 510

```

1237 GAATTGGACGAGAAGACCTCGC 1259
 11 || ||||| |||
 510 LysArgGlyArgArgArgAspArg 517

seq_name: /cgn2_6/ptodata1/iaa/6B_COMB.pep:US-09-413-814-78

```
seq_documentation_block:
```

; Sequence 78, Application US/09413814

; Patent No. 622506

; GENERAL INFORMATION:

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, Co

APPLICANT: Beyer, Stefan

APPLICANT: Beyer, Stefan
APPLICANT: Bloeker, Helmut

APPLICANT: Bloecker, Hell
APPLICANT: Brandt, Petra

APPLICANT: Brandt, Petra

APPLICANT: Cino, Paul M

APPLICANT: Dougherty, Brian A

APPLICANT: Goldberg, Steven L

APPLICANT: Hofle, Gerhard

APPLICANT: Mueller, Joachim

```

; APPLICANT: Mueller, JOACHIM
; APPLICANT: Reichenbach, Hans

```

APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic

TITLE OF INVENTION: heteropolynucleotide compositions and methods for determining DNA sequences for enzymatic cleavage

;; TITLE OF INVENTION: heteropolysketide comp
AND REFERENCE: PCT/US 99/23535

FILE REFERENCE: PCT/US 99/23535

CURRENT APPLICATION NUMBER: US/09/413,814

CURRENT FILING DATE: 1999-10-07

; EARLIER APPLICATION NUMBER:

EARLIER AFF
; EARLIER FIL

; NUMBER OF SE

NUMBER OF SE
SOFTWARE: Pa

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78

```

; SEQ ID NO 78
LENGTH: 887

ORGANISM: Sorangium cellulosum
US-09-413-814-78

```

        alignment_block:

```

Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

```

140 AAGCGCATGCCGTCAAAAAGGCCAAGTGCCTGTTTGAGACGACCAAAAAGNAT 189
      ::::: ||| ::::: ::::: ||| ::::: :::::
312 ArgIyArGPrOPrOAlaIarGserSerAlaCyArGArGThrArGserSe 328
      ||| ::::: ||| ::::: ||| ::::: ||| :::::
130 CCGGGCGTGGTCTTACCGCGCCGCGNCTTTCAGGCAAAATCGCGGCATCCA 239
      ||| ||| ||| ||| ||| ||| |||
338 tGlyArGtGlyThrGlySerArGArGAlaAlaIaIaPrOPrOPrOProSer. 344
      ||| ||| ||| ||| ||| ||| |||
240 TCGCGCGCAAAAGCCGCTACTCAGTCGGCTCGATTCGCGTTGAGGCA 289
      ||| ||| ::::: ||| ||| ||| ||| :::::
345 ArGArGArGPrOPrThrArGPrOserArGPrOAlaYArGThrPrOserArGAr 361
      ||| ||| ||| ||| ||| ||| ||| |||
290 ACGGCAAAATCGATGTTCGAAAGCCTACGCGCCGCGAAGCGTTGGCAACTTA 339
      ||| ||| ::::: ||| ||| ||| ||| |||
361 yArGArGserPrOAlaArGThrPrOgLyPrOArGser.....P 374
      ||| ||| ||| ||| ||| ||| ||| |||
340 ACGGCGGAGGAANTNNGNCATCTGATTCACATCCGGTTTGCGACTGC 389
      ||| ||| ||| ||| ||| ||| ||| |||
374 rOArGArG.....ArGPrOserPrOser 381
      ||| ||| ||| ||| ||| ||| ||| |||
390 GGTGCGTANCCGTCGTTACGCAAAATCCCTCGCTCGA..... 428
      ||| ::::: ||| ::::: ||| ::::: ||| ||| ||| |||
382 AlaArGserPrOAlaIeTrpGluArGPrOPrOArGArGPrOArGAspAr 358
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 .....TGCGAGCCGCTGGCCACTTCTGCGAATGGCAT. 461

```

398 gatgptatgtagtgleuleuartrgprohlaartrgtagtgleulaatargtaaspr 415
462 ||| :|||:||||| :||| :|||
415 roglasprvalproasprroargalalalartrargtrgalprohlaala 431
480 GCGACACCCCTGTGGTTGTGATCCAAAGAGC.....CG 511
448 ||| |||:||||| |||:||||| :||| :|||
432 gylalaleuproglylalaasprglyargglyalargglyprogllyargtr 448
512 NCGANGATTTTCAGACGANGNGTGTGTGATTGACCGGTTTACCGAGCGT 561
448 |||:| :||| :|||:||||| ||| |||:|||||
448 gargalalarproargtrargglyalalalalaartrgalproproasprgalaa 465
562 AAATCCATGTGTGTAAAGCAGCTGGGGCAGACGTTGCCGCTGAAAAATGC 611
465 sppropro..... 467
612 TGCCAAACATCGAACAACATGTAATTCGGCGGCCGATCCGGCGGTTTGA 661
468 |||:||||| |||:||||| :||| :|||
662 GTGCGACCGCACATTCATTCTTACGTGACCGCGGTGGTGAACAAACCGT 711
476 :||| :||| :|||:||||| |||:||||| :||| :|||
476 gylalalarproalavalalartrargtrgsealalartrgalglyelprogly 493
712 TGACCATTCATTAATCAAGATGTAATTCGATCCGACGTTTGTTCGAC 761
493 :|||:||||| :||| :||| :||| :||| :||| :||| :|||
493 alglahlisproval..... 497
762 AGCGCGTGTGAAACACCGACCGCGGTGATGCTTGGTGTTCTCAATCA 811
498 |||:||||| |||:||||| :||| :|||
812 ACAAAACACGCGCTGTGGGTACCGGCTTTGGGTGGCAAGTATCGCAATT 861
510 |||:||||| :||| :||| :|||:||||| :||| :|||
510 aoproproglinprovalalartrargalalalarproargglyalalalaaspr 527
862 ACTGCGCGCGAATTGGTTGTGACCCAGACAAACCGCGGTGATTCGGT 911
527 sptalalnartrglngralargalavalpro..... 537
912 ATTGAACGCGCGGATTCACACAGGGCGGCGACGATTATTGGAGCGTACC 961
538 |||:||||| |||:||||| :||| :|||
962 ACATCAGATTTCCGTTATCGAAGAGCGCGACGAAGACGTGTTGG. 1010
548 ||| ||| ||| |||:||||| :||| :|||
548 o.....Aspralartrg...Proarglalarproargglyargglyproarg 562
1011CTGGTTGCGCGCGCAGCGCGGCAAAATCTCCATCAGCCGTRAC 10522
562 rglglyleuleuglyalalartrproargglyelulproarglavalalarpro 578
1053 GACCCCTCGGCATTTCTCTGAAAAACAACTCTTCAAGTTCAAGCAGACGCG 11022
579 gluproarglalarpro..... 584
1103 TCAAGGTGGCGACCGCGCATGTGTCCGATTGTACTAGAGGCGGA 11522
585 |||:||||| |||:||||| :||| :|||
1153 ATGCGCGTAGACATTCGCGCTTACCGCTGCTTGGTTCGCGATTATTCGT... 11992
594 rgalalalartrghislartrgalalartrgalalarproargtrgalalartrgleu 610
1200CGCGGATACGACGACGCGCGACAGCATTTGGGTTCGTTGGAATTGG 1244
611 ||| ||| |||:||||| |||:||||| :||| :|||
611 Valglylartrgleualartrgalalartrgalalartrgalalartrgleu... 626
1244 ACGAAGAAAGACCTGCTTGTGACGCTTCGTCGCCGGGCAAAATACGA 1292
||| ||| ||| |||:||||| |||:||||| :||| :|||

```

637 .ArgAlaGlyProAlaIleProAlaIleAlaGlyValaProGlyValaValArg 642
seq_name: /cgcn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-997-897-2
seq_documentation_block:
Sequence 2, Application US/08997897C
Patent No. 6114514
GENERAL INFORMATION:
APPLICANT: SRIVASTAVA, RANJANA
APPLICANT: KUMAR, DEEPAK
APPLICANT: SRIVASTAVA, BRAHM SHANKER
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT
FILE REFERENCE: u011469-7
CURRENT APPLICATION NUMBER: US/08/997,897C
CURRENT FILING DATE: 1997-12-24
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 430
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: UNSURE
LOCATION: (4)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (6)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (20)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (29)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (54)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (64)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (69)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (89)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (99)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (114)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (119)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (129)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (159)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (169)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (182)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (185)
FEATURE:
NAME/KEY: UNSURE
LOCATION: (219)
FEATURE:

```

1	NAME/KEY:	UNSURE
2	LOCATION:	(259)
3	FEATURE:	
4	NAME/KEY:	UNSURE
5	LOCATION:	(269)
6	FEATURE:	
7	NAME/KEY:	UNSURE
8	LOCATION:	(291)
9	FEATURE:	
10	NAME/KEY:	UNSURE
11	LOCATION:	(323)
12	FEATURE:	
13	NAME/KEY:	UNSURE
14	LOCATION:	(339)
15	FEATURE:	
16	NAME/KEY:	UNSURE
17	LOCATION:	(349)
18	FEATURE:	
19	NAME/KEY:	UNSURE
20	LOCATION:	(356)
21	FEATURE:	
22	NAME/KEY:	UNSURE
23	LOCATION:	(366)
24	FEATURE:	
25	NAME/KEY:	UNSURE
26	LOCATION:	(410)
27	FEATURE:	
28	NAME/KEY:	UNSURE
29	LOCATION:	(421)

```
alignment_scores:
  Quality: 112.00
  Ratio: 0.615
  Percent Similarity: 41.270
  Length: 441
  Gaps: 24
  Percent Identity: 23.583
```

```
alignment_block:
```

```

US-09-303-518D-127 x US-08-997-897-2  ..
Align seg 1/1  to: US-08-997-897-2  from: 1  to: 430

```

```

44 GACGGGAGCAAGTCATTTATATACGGGGCCGGTCATTA.....CCGA 84
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
34 Asptdgluileupspgluglusergltprproglytyrargpropropar 50
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
85 GTCGCGTTGCTTGCCAGCAAAATGTGCCGGTATAGCCGCCCTGATGAAGT 134
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
50 gserhnicys***Ser.Argprocyarg.....Lysarg 61
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
135 CAAGGAAGGGATGCCGTCAAAAAGGCC.AAGTCC.....169
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
62 Argtyr***Argcysargasp***Argserargcystprcysnhsprogl 78
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
170 .....TGTTAAAGCAAAAAGNATCGGGCGCGT 200
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
78 ytyrsercysargtyrtyrprcysnhsproarg***SerCysargcystprc 95
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
201 GTTTACCGCGCCCGTTTCAGGCAAAATGCGCCCATCCATCGCGCGCAA 250
   |||||||:::|||||:::|||||:::|||||:::|||||:::
95 ys-tyrtyrarg***Sertrargtyrcyscyssercysprcscysargse 111
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
251 AGCGCGACTTCAGTCGGCTCGATGGCGGTGAGGAAGCAAGCAATC 300
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
111 rProCys***Ser.....ArgArgarg***G 120
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
301 GAETTGGAACGCTACGCGCCCGCAAGCGTTGGCAACTTAAGCGCGANGA 350
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
120 lycsPrOCyscyssercysglnhns***glyCysargtyrCysargtyr 136
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
351 ANTNMNNNCATTCGATCCATCAATCGGGTTGTGGACGCGCGCTCGCAGNC 400
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
137 ProGlySerargTyrrProSerSerarg.....Cys.....pr 147
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```

401 GTCGCTTCAGCAAAATCCCTGCTGCATCCGACGCCGTTGCCATCTTC 450
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 oserleuarg.....CysArgArgPheArgCysPProArg***ArgC 161
451 GTCATATCGCATGACACCAATCCGCTGCGGACAGCCCTGTGTGTGGAT 500
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 ysGlnArgTgtTyrPCysProAsn***ThrGlyArgCysCysArgCysPro 177
501 CAAGAAGAGCCGNCAGATTCAGACAGANTGTCGTATGAGCCGTT 550
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 SerSerSerArg***GlnTyr...SerProAlaGlyCysArgArgGln 193
551 TGACCGAGCGGTAATTCATGTGTGTAGGACAGCGCGCGAGAGTCCG 600
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 rAlaArgCys...ArgCysCysCysCysArgCysTyrPArgThrArgCysC 209
601 TCTGAATAATGCTGCCAA...CATCGAAACACATGATTCGCGCGCCGCA 647
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 ys.....CysCysArgCysTyrPArgInsLeuGly***SerArgProArg 223
648 TCCGCGCGGTTTGTAGTGACGACCATTCATTCATGAGCCGCGGTG 697
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 SerArg.....SerArgArgCys.....SerArgArgCys 229
698 CAACAACAAACCGTTTGGACCATCATATTCAGATGTATTCGATCGGA 747
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 sSerArgArgArgPheGlnAsnArgCysCysArgSerArgGlyPheArgI 246
748 CGTTTGTTCACACAGCCG.....TCTGACACAC.....* 776
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 leArgCysCysSerPheProGlyPheArgAsnArgHis***IleLeuArg 262
777 .....CGAGCCGCGTATG 790
263 CysPheHisCysArgTyr**SerCysArgArgCysArgCysArgArgCys 279
791 CTTTGGGAGGTTCTCAAGTCAA...CAACACACCGCTTGGCTACCGTT 837
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 sPheGlyCysArgGlyCysArgCysGlnGlyCys***SerHisArgArgP 296
838 TTGGGTGGCAAGATTCGCAANT...TACGCGGCGCAATTCGTTGTCGCG 884
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 heArgCysArgGlyCysCysAsnCysArgCysTyrArgCys..... 309
885 AGACACCGCGTATTCGCGTTCGATTCAGCGCGCGATTCACACAG 934
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 .....ArgGlyCysSerAr 314
935 GCGCGGACGATTTATTTGGG.....ACGC 957
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 gatgProGlyLeuProGlyArgAsp***ArgProValGlyHisArgIysI 331
958 TACCACATAGATTCGTTACGAAAGAGCGCCGACCAAGAGAGCGTT 1007
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
331 leProThrCysCysPheArgCys***ArgSerProArgSerArgProAla 347
1008 CGGCTG.....GGTTCGCGCGACCGCGGACAAATCTCCATCAGCGC 1048
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 leu***TrpProProGlySerCys***ThrAsnProIleArgCysCysPr 364
1049 GTACGACCGCTCGGCAATTCCTGAAAAACAAACTTCAGATTCACAGCA 1098
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 oSer***SerArgProIleProAlaArgProArgLeuPro.....GlyA 379
1099 GCGGTCAACGGTGGCGACCGCGCCAGTGGCCGATTCGACTAGACGGC 1148
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
379 rgSerTyrArgTyrProProThrIlySerGly.....ArgSer 391
1149 CGTATGCGCGGCTAGACATCTCGCTACCGCTTTCGCGGATTTAATCG 1198
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392 GlnAsnCysTyrPHisArg.....SerSerGlySerAr 402

```

```

1199 TCGCGCATACGACAGCGC 1217
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 gThrArgTyrArgHisArg 408
seq_name: /cgn2_6/plodata/1/laa/5B_COMB.pep:US-09-156-836B-2
seq_documentation_block:
; Sequence 2, Application US/09156836B
; Patent No. 6242585
; GENERAL INFORMATION:
; APPLICANT: Srivastava, Ranjana
; APPLICANT: Kumar, Deepak
; APPLICANT: Srivastava, Brahm Shanker
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT
; FILE REFERENCE: U 011876-4
; CURRENT APPLICATION NUMBER: US/09/156,836B
; CURRENT FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 08/997,897
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (4)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (6)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (20)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (29)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (54)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (64)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (69)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (89)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (99)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (114)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (119)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (129)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (159)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (169)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (182)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (185)

```

OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (219)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (259)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (269)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (291)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (323)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (339)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (349)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (356)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (366)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (410)
OTHER INFORMATION: amino acid has not been identified
NAME/KEY: UNSURE
LOCATION: (421)
OTHER INFORMATION: amino acid has not been identified
US-09-156-836B-2

alignment_scores:

Quality: 110.00 Length: 446
Ratio: 0.625 Gaps: 23
Percent Similarity: 39.462 Percent Identity: 23.767

alignment_block:

US-09-303-518D-127 x US-09-156-836B-2 ..

Align seg 1/1 to: US-09-156-836B-2 from: 1 to: 430

44 GACGAGACGATGATTTATGACGGCGCGTCA...CCGA 84
||||| : : : : : : : : : : : : : : : : : :
34 Asparaginilaspriylserglytyrargproproar 50
85 GTCCGCTGCTGGCGAAGATATGCGGTATGCGCCCTNGATGAAGT 134
:
50 gserhisCys***Ser ArgProCysArg.....LysArg 61
135 CAAGAGAGCGATGCGTCAAAAAGGCC.AAGTGC..... 169
:
62 ArgTyr***ArgCysArgasp***ArgSerArgCysTyrCysHisProgl 78
170TGTGGAAGCAAAAAGNATCCGGCGGT 200
78 YTYrSerCysArgTyrTyrCysHisProArg***SerCysArgCysTyrC 95
201 GTTTACCGCGCGNCTTTCAGGCAAAATGCCGCATCCATCGCGCGAAA 250
||| : : : : : : : : : : : : : : : : : :
95 Ys-TyArgArg***SerArgArgCysSerCysProCysCysArgse 111
251 AGCGCTACTTCATGTCGTCGTGATGCGGTGAAGCAAGCGAAGATC 300
:
111 rProCys***Ser.....ArgArgArg***G 120
301 GAGTTGAAAGCGTACCGCGCGGAGCGTTGCAAACTTAAGCGCGANGA 350
:
|||

120 lYcysProCysCysSerCysGlnHis***GlyCysArgTyrCysArgTyr 136
351 ANTNNNGNCAANTCMTATCCATCCGCTTTGTGGACCTGCGTANCC 400
:
137 ProGlySerArgTyrProSerSerArg.....Cys.....Pr 147
401 GTCCGCTTTCAGCAAAATCCCTGCGTGCATGCCGCGTCCGCACTTC 450
||||| : : : : : : : : : : : : : : : : : :
147 oSerLeuArg.....CysArgArgPheArgCysProArg***ArgC 161
451 GTCAATGCGATGACACCAATCCGCTGCGGAGACCCCTGTGTGAT 500
||| : : : : : : : : : : : : : : : : : :
161 YsGlnArgTyrTyrCysProasn***ThrGlyArgCysArgCysPro 177
501 CAAGAAGCGCGNCGANATTTGACAGCANGTNGCTGATGAGCGCTT 550
:
178 SerSerSerArg***GlnTyr....SerProAlaGlyCysArgArgTh 193
551 TGACCGAGCGTAAATCCATGTCGTGAAGCAGCTCG..... 587
193 lAlaArgCys...ArgCysCysCysArgCysTyrPargThArgCysC 209
587 587
209 YsCysCysArgCysTyrPglInserLeuGly***SerArgProArgSerArg 225
588CGCAGAGCTGCGCTGTAATAATGTCGCAACATCG 622
226 SerArgArgCysSerArgArgArgPheGlnAsnArgCysArgSerArg 242
623 AAACACATGATTCGCGCGCGCCATCCGCGCTTGTAGTGCGACGAC 672
242 gGlyPheArgIleArgCysCysSerPheProGlyhe.....A 255
673 ATTCAATTCATTTGAGCGCGTGCATGTCGCAAAACCGTTTGACATCA 722
255 rgAsnArgHis***IleLeuArgCysPheHisCysArgTyr....Ser 270
723 TTATCAAGATGTAATTTGCCATCGACGTTTGTTCGAACAGCGCGCTGA 772
271 CysArgArgCysArgCysProArgCysPheGlyCys..... 282
773 ACACGAGCGCGTGAATGCTTTGGTGGTCTCAAGTCACAAACACGCG 822
283 ...ArgGlyCysArgCysGlnGlyCys***SerHisArg..... 294
823 CTCCTGCGTACCGTTTGGTGCAGAAAGTATCGCAAT...TACTCGCG 869
295ArgPheArgCysArgGlnCysCysAsnCysArgCysTyr 307
870 GCAATTTGTTGACCGCAGCAACCGCGTGAATTCGGGTTGATTCGACG 919
308 ArgCys..... 309
920 CGCGATTCACAAAGCGCGCACATTAATTTGG..... 953
310 ArgGlnCysSerArgArgProGlyLeuProGlyArgasp***ArgPro 326
954ACGCTACACAAATCAATGATTCGCTTATGAGAAAGCGG 992
326 alGlnHisArgIleProHisCysPheArgCys***ArgSerPro 342
993 CAGCAAGAGCTGTCGCTG.....GGTTCGCGCGAGCGGAGCA 1033
343 ArgSerArgProAlaLeu***TyrProProGlySerCys***ThAsnPr 359
1034 AATATTCATTCACGCTACGACCTTCGCGCATTCCTGTAACAAACATC 1083
||| : : : : : : : : : : : : : : : : : :
359 oIleArgCysCysProSer***SerArgProIleProAlaArgProArg 376
1084 TTCAGTTTCACGAGCGGTCAACGCTGACGCGCGCGCATGTCGCGAT 1133
||| : : : : : : : : : : : : : : : : : :
376 eupro.....GlyArgSerTyrArgTyrProProHisSerGly... 389

1134 TGGTACTACGAGCGCGTATCCGCTAGACATCGCTGCTACCTGCTTT 1183
||||| : : : : :
390 ArgSerGlnAsnGlyTrpHisArg..... 397
1184 TCGCGGATTTAATGCTCGCGCATACCGACAGCCG 1217
: : : : :
398 .SerSerGlySerArgThrArgTyrArgHisArg 408
seq_name: /cgn2_6/prodata/1/aa/backup/seq.5223423-4
seq_documentation_block:
; Patent No. 5223423
; APPLICANT: FRANCHINI, GENOVEPPA, WONG-STAL, FLOSSIE;
; GALLO, ROBERT
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/331,212
; FILING DATE: 03-31-1989
; SEQ ID NO: 4:
; LENGTH: 3080
5223423-4
alignment_scores:
quality: 104.50 length: 433
ratio: 0.562 gaps: 25
Percent Similarity: 42.956 Percent Identity: 21.940
alignment_block:
US-09-303-518D-127 x 5223423-4 ..
Align seg 1/1 to: 5223423-4 from: 1 to: 3080
195 CGTGGTGTTCACCGCCGCTTCAGCAAAATCCGCCCATCCAGCGC 244
||| : : : : :
256 ArgHisLeuValLeuAlaArgArgGluSerGluArgTyrGlySer 272
: : : : :
245 GCGAAACCGCGT...ACTTCAGTCGCT...CGTATTCGCTTGA 285
: : : : :
272 slyssnsrArgGluThrSerSerGlyArgAsnArgAsnGly 289
: : : : :
286 GCGACGACGAATCGAGTTCGACGCTACGCGCCGACGCTTGGCAA 335
: : : : :
289 snalalysTyrLysThrAsnSerThrThrArgGluArg.....GlyLys 303
336 CTTAGCCGCGGANGAANTNNNGCAATCTGATCCATCCGCTTGTGA 385
||| : : : : :
304 LeuProArgAlaThrAsnArgArgGln..... 312
386 CTGCGCTGCGTANCCGCTCCGTCAGCAAAATCCCTGCCGTCGATGCCGAG 435
||||| : : : : :
313LeuGlyProSerAlaAlaGluSerProAsnProLysGlyLeuG 327
436 CC...GTTGCCCATCTTCGTCGTAATGAGTGCACCAATCCGCTGNC... 479
: : : : :
327 LysValSerArgGlyGlyLysValArgGlyArgSerSerAlaGlyLe 343
480GG 481
344 SerGlyThrLeuArgArgLeuHisAlaLeuTyrSerAsnAlaLeuGly 360
482 CAGACCTGTGGTGTTCATCAAGAGCGGANGATTCAGACGAGGT 531
||||| : : : : :
360 yArgProSerSerSerAspAlaAsnGlnArgAsnTyrArgSerSer 377
532 NTGCGGATTGAGCCGTTGACCGAGCGTAAATCATGTGTGAAGC 581
||| : : : : :
377 rgleugly..... 379
582 AGCTGGCGCAGAGCTGCCGCTCTGAATAATGCTGCAACATCGAAACATG 631

380CysThrThrSerAsnThr... 385
632 AATTCGCGCGCCGATCCGCGGCTTGTAGT.....GCGACGACATTT 675
||||| : : : : :
386ArgProLeuThr.SerGlyAlaAlaGlnArgSerThrArgGly 399
676 CATTCATTTAGCCGCGTGCATCAACAAACCGTTTGACATCAATTA 725
||| : : : : :
400 HisSerArgAspAsnLysHisSerArgGlyThr..... 410
726 TCAAGATGATATTCGATCGAGCTTTGTTTGCACAGCCGCTGACAA 775
||| : : : : :
411AspArgMetAspValAlaArgLysSerCysThrSerArgLysHis 426
776 CCGAG.....CGGCTGATTCCTTG 795
426 euGluMetAspProAspArgThrAlaGluValCysGlnAsnValGlnSer 442
796 GGTGCTTCAGAGTCACACAAACCGCTTCGCTACCGTT..... 837
: : : : :
443 AsnGlnHisSerArgHisLysThrArgThrLysArgValAlaProLysLe 459
838TTGGTGCAGAAAGTATCGCAATATTCGCGG 868
459 uCysGlyLeuLeuGlnLysLeuLysGlyArgThrAspArgCysSerSer 476
869 GCGATTGCTGTGACGACGAC.....AACCGCTGATT 900
||||| : : : : :
476 LglLleuAspAspProAspAlaAlaSerAlaLleuGluProArgLeuVal 492
901 TCCGCTTCGATATTGAACGCGCGATTCACACAAAGCGCGACGAT..... 945
493 Ser.....ThrLysGlyThrArgAspLysLeu 501
946 .TATTTGGAGCGCTACCAATTCGATTTCGTTATCGAAGAGCGCGCA 994
||||| : : : : :
501 rTyrLeuArgArgAspAlaAsnArgLeuSer.....Arg 513
995 GCAAGAGCGTTCGCTGCGCTGCGCGCGACGCG.....GACAA 1034
513 spArg.....TrpThrArgProGluGlnThrAsnGlyArg 525
1035 ATACTCATACAGCGCTACAGACCGCTGCGCATTTCT..... 1070
526 SerLeuLysArgGlyHisAlaThrSerProTyrProLecYssSerSer 542
||| : : : : :
1071GAAAAA...CAAACTTCAAGTTTCAGACAGCGCT..... 1103
542 oThrGluLysGlyAsnValLeuGlnLeuThrLysGlyArgAlaLeuGly 559
1104CAACGGTGG.....CGA 1115
559 ysrThrMetProSerAlaLysThrGlyLeuLeuGlnMetThrProLysValArg 575
1116 CCGCGCCATGTT.....GCCGATTGGTACTTACGAGCG..... 1148
: : : : :
576 ThrHisHisGlyLysLeuProArgThrGlyTyrPheArgAlaTyrPth 592
1149CGTATGCCCGCTAGACATCTGCGCTACCTGCTTGGCGCAT 1191
: : : : :
592 rMetGlyLysGluAlaProGlnLeuProArgGlyProLysPheAlaGly 609
1192 TTAATCGTCGCGGATACGACAGCGCGCAAGATGGGTTGCTTGA 1238
||| : : : : :
609 laAsnThrAsnSerThrProAsnGlySerSerGlyProThrGly 624
seq_name: /cgn2_6/prodata/1/aa/5A_COMB.pap:US-08-046-585-5
seq_documentation_block:
; Sequence 5, Application US/08046585
; Patent No. 5453362
; GENERAL INFORMATION:

```

1  APPLICANT: Lamarco, Kelly
2  APPLICANT: Wilson, Angus
3  APPLICANT: Herr, Winship
4  TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN
5  TITLE OF INVENTION: HOST CELL FACTOR
6  NUMBER OF SEQUENCES: 15
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
9  STREET: 4 Embarcadero Center, Suite 3400
10 CITY: San Francisco
11 STATE: CA
12 COUNTRY: USA
13 ZIP: 94111-4187
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/046,585
22 FILING DATE: 12-APR-1993
23 CLASSIFICATION: 433
24
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Osman, Richard A
27 REGISTRATION NUMBER: 36,627
28 REFERENCE/DOCKET NUMBER: A-57503-1/RAO
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (415) 781-1989
31 TELEFAX: (415) 398-3249
32
33 INFORMATION FOR SEQ ID NO: 5:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 2035 amino acids
36 TYPE: amino acid
37 STRANDEDNESS: single
38 TOPOLOGY: linear
39
40 MOLECULE TYPE: peptide
41
42 US-08-046-585-5

```

alignment_scores:	
Quality:	100.50
Ratio:	0.333
Percent Similarity:	40.379
alignment_block:	
length:	634
Gaps:	32
Percent Identity:	20.662

US-09-303-518D-127/rev x US-08-046-585-5 ..
Align seg 1/1 to: US-08-046-585-5 from: 1 to: 2035

[illegible]

1145TCTAGTAGCAATCGGACCCATGCGCGGTGCGCAC 1108
843 G|Y|T|H|T|L|E|U|A|G|T|H|V|A|L|P|R|O|T|E|I|L|Y|L|A|G|L|E|U|A|L|H|T|H|P| 11
1107 G|T|T|G|A|C|G|.....G|C|G|T|G|C|A|A|C|T|T|G|A|A|G|T|T|G|T 1076
859 O|A|L|T|H|V|A|L|S|E|R|A|L|V|A|L|Y|S|P|R|O|L|A|V|A|L|T|H|T|H|L|E|U|..... 872
1075 T|T|T|C|A|G|A|A|T|G|C|C|G|C|G|G|G|T|G|C|C|A|A|T|A|T|C|G|G|C|G|C|T|T|G|T|A|A|C|G|G|C|G 1047
873V|A|L|L|Y|S|G|L|H|T|H|T|H|G|L|V|A|L|T|H|T|H| 882
1046G|T|G|A|T|G|A|G|A|T|A|T|T|G|T|C|G|C|G|G|C|G|C|G|C|G 1018
883 L|E|U|G|L|Y|T|H|V|A|L|H|T|H|G|L|Y|T|H|V|A|L|S|E|T|H|S|E|R|L|E|U|A|L|A|G|L|Y|L| 899
1017 A|A|C|C|A|C|C|C|G|A|C|G|C|T|T|G|C|T|G|C|G|G|C|C|T|T|G|A|T|A|C|G|A|A|T|C|T 968
899 Y|H|S|E|T|H|S|E|R|A|L|S|E|U|..... 906
967 G|A|T|T|G|T|G|A|G|C|T|C|C|A|A|T|A|T|C|G|G|C|G|C|T|T|G|T|A|A|C|G|G|C|G 918
907A|A|H|H|P| 909
917 T|T|C|A|A|T|C|G|A|C|C|G|A|A|T|C|A|C|G|G|T|T|G|T|G|C|T|G|C|A|C|C|A|A|T|T|C|G|C 868
910 I|L|E|T|H|T|H|L|E|U|G|L|Y|H|T|L|E|A|L|H|T|H|S|E|S|E|G|I|N|A|L|I|L|E|A|S|P|R 926
867 C|G|C|A|G|A|A|T|T|G|C|G|A|C|T|T|T|G|C|A|C|C|A|A|C|G|G|T|A|C|G|A|G|A|G|C|G|T|G 818
926 O|H|R|A|L|L|E|T|H|V|A|L|S|E|R|A|L|A|G|L|T|H|T|H|L|E|U|H|T|H|A|L|A|G|L|Y|G 943
817 G|T|T|G|.....T|T|G|A|C|T|T|A|G|A|C|A|.....C|C|C|A|A|G|C|A 789
943 I|Y|L|E|U|T|H|T|H|P|R|O|T|H|L|E|H|T|H|E|T|G|I|N|P|R|O|V|A|L|S|E|G|I|N|P|R|O|T|H|C|I|N 959
788 A|T|C|A|G|.....C|G|C|G|G|G|T|T|C|A|G|C|G|C|G|T|T|..... 759
960 V|A|L|T|H|L|E|U|L|E|T|H|A|L|P|R|O|S|E|G|I|Y|A|L|G|I|N|A|G|I|N|P|R|O|V|A|L|H|I|S|A|S 976
758G|C|A|A|C|A|A|A|C|G|T|C|C|G|A|T|G|G|C|A|A 737
976 P|L|E|U|P|R|O|V|A|L|S|E|T|L|E|U|A|L|S|E|T|P|R|O|T|H|T|H|G|L|I|N|P|R|O|T|H|A|L|A|T 993
736 T|T|A|C|A|T|C|T|G|A|T|A|T|A|T|G|T|C|C|A|A|C|G|T|..... 706
993 H|R|V|A|L|T|H|L|E|A|S|P|E|R|G|L|Y|G|I|N|A|S|P|R|O|G|I|Y|T|H|V 1009
705T|T|T|G|T|T|G|C|A|C|G|C|A|.....C|G|G|T|C|A|A|A|T|G|A|A|T|G|G|G 670
1009 A|L|H|R|L|E|U|V|A|L|Y|S|S|E|R|A|S|P|R|O|C|Y|S|E|L|U|H|T|H|S|G|L|U|T|H|T|H|G|L|Y|T|H 1025
669 C|G|T|G|C|C|T|C|A|A|C|G|C|G|C|G|C|G|C|G|C|..... 640
1026 T|H|A|S|T|H|L|A|L|H|T|H|T|H|H|V|A|L|A|A|A|S|N|L|E|U|G|L|Y|H|I|S|P|R|O|G|I 1042
639G|C|C|A|A|T|C|A|T|G|T|..... 625
1042 N|P|R|O|T|H|G|I|N|V|A|L|G|I|N|P|H|E|V|A|L|C|Y|S|A|S|P|R|A|T|G|I|N|G|I|U|A|L|A|A|L|S|E|L 1059
624T|T|G|A|T|G|T|G|C|A|G|A|C|A|T|T|T|C|A|G|C|G|C|G|C|A|C 595
1059 E|U|V|A|L|T|H|S|E|T|H|V|A|L|G|I|N|G|I|N|A|S|N|E|L|Y|S|E|V|A|V|A|L|A|T|G|..... 1073
594 G|T|T|G|C|G|C|A|C|.....T|G|C|T|T|A|C|A|C|A|C|A|T|G|A|T|T|T|T|A|C|G 559
1074 V|A|L|Y|S|S|E|R|A|S|P|R|O|C|Y|S|E|L|U|H|T|H|S|G|L|U|T|H|T|H|G|L|Y|T|H|T|H|A|S|T|H 1090
558C|T|G|G|T|C|A|A|.....C|T|G|G|T|C|A|A|..... 550
1090 R|A|L|A|T|H|T|H|A|L|H|T|H|S|E|R|A|S|A|M|E|T|A|L|A|G|L|Y|G|I|N|H|I|S|G|L|Y|C|Y|S|S|E|R|A|S|P 1107


```

COUNTRY : USA
ZIP : 94111-4187
COMPUTER READABLE FORM :
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,703
FILING DATE: 24-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-2/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-393-703-5

alignment_scores:
Quality: 100.50 Length: 634
Ratio: 0.393 Gaps: 32
Percent Similarity: 40.379 Percent Identity: 20.662

alignment_block:
US-09-303-518D-127/rev x US-08-393-703-5 ..

Align seg 1/1 to: US-08-393-703-5 from: 1 to: 2035

1343 CAGCGCTCCGTCGCACANGGTGGTCCACACACTTGACGCAACAGCGGCCNTA 1294
      ::::::::::::::::::::|::::::::::::|::|
747 LysProThrIleLeuGlyIleSerSerVal.....SerProse 759
1293 TTGCGATTTGCCGCGGACGACGAGAGCTGCACAAGAAGGCTTTCTTCGT 1244
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
759 rThrTrHlvsProGlYthrThrThrIleIleLysThrllePrometsera 776
1243 CCATTCGACGACCAACCATGCTTGCGCGCTGCGGATGCGGACGACATT 1194
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
776 lallelelrhGlnAlaGlyAlarThrGlyValThrserSerProGlylle 792
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1193 AAATCG..... 1188
      ||||||
793 LysSerProIleThrIleIleThrThrLysValMetThrSergIyThrgl 809
1187 .....CCAAAGACGAGGTAGGACAGA 1166
      |||
809 yalaProAlaLysIleIleThrAlaValProLysIleAlarThrSylsg 826
1165 TGCTTAGCGGACATPACGCG..... 1146
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
826 LygInGInGlyValThrGlnValValLeuLysGlyAlaProGlyLnPro 842
1145 .....TCGTAAGTACCATGGCACCATGGCGCGGTGGCCACC 1108
      ||||||::|::|::|::|::|::|::|::|::|::|::|::|::|
843 GlyThrIleLeuArgThrValPrometolYglYalArGleUvalThnPr 859
1107 GTTGACG.....GCGTGCGTAACCTGAAGGTTTGT 1076
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
859 ovalThrValSerAlaValLysProAlaValAlThrThreU..... 872
1075 TTTTCAGCAAAATGCGCGAGGCTGCTACGC..... 1047
      ||||||::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

873 .....ValValysglyThrThrThrThrThr 882
1046 .....GTGATGAGTATTTGTCGGCTCGGGCC 1018
883 LeuGlyThrValThrGlyThrValSerThrSerLeuAlaGlyAlaGly 899
1017 AACCCAGCCAGCAGCTTTGGCTCGGGCTTTCCGATACGGAATCT 968
899 yHisSerThrSerAlaSerLeu..... 906
967 GATTGTGTAGCGTCCAAATATATGTCGGCCTTGTAATCGCGCC 918
907 .....AlaThrPro 909
917 TTCAATACCGAACCAGAAATACCGGTTGTCTGCTCAACCAATTCGCC 868
910 IleThrThrLeuGlyThrIleAlaThrLeuSerSerGlnValIleAsnPr 926
867 CGCAGTAATTTGGCATNCTTCCGACCCAAACGATGCGCAAGCGCGT 818
926 oThrAlaIleThrValSerAlaAlaGlnThrThrLeuThrAlaAlaGly 943
817 GTTGTG.....TTGACTTGAGACCA.....GCCAAGCA 789
943 LysLeuThrThrProThrIleThrMetGlnProValSerGlnProThrGln 959
788 ATCAGC.....CGCTCGGTGTTCAGACGGCGCTGT..... 759
960 ValThrLeuIleThrAlaProSerGlyValGlnAlaGlnProValHis 976
758 .....GCAACCAACGTCGAGGCA 737
976 PheProValSerIleLeuAlaSerProThrThrGlnGlnProThrAla 993
736 TTCAATCTGATAATGATGTCACAAACGT..... 706
993 hr..ValThrIleAlaAspSerGlyGlnGlyAspValGlnProGlyThr 1009
705 .....TTTGTTCACCGAC.....CGGCTCAATGAATGATGTG 670
1009 alThrLeuValCysSerAsnProProCysGluThrHisGluThrGlyThr 1025
669 CGTGCACCTCAACCGCGCGGATGCGCGCC..... 640
1026 ThrAsnThrAlaThrThrThrValValAlaAsnLeuGlyGlyHisProG 1042
639 .....GCCGAATCATGTGT..... 625
1042 nProThrGlnValGlnPheValCysAspArgGlnGlnAlaAlaSerL 1059
624 .....TTGATGTTGGCAGCATTTCAGACGGCAC 595
1059 euValThrSerThrValGlyGlnGlnAsnGlySerValAlaArg..... 1073
594 GTCTGGCGCAGC.....TGCCTACACACATGATTTAG..... 559
1074 ValCysSerAsnProProCysGluThrHisGluThrGlyThrAsnTh 1090
558 .....CTCGTCAA..... 550
1090 rAlaThrThrAlaThrSerAsnMetAlaGlyGlnHisGlySerSerSnp 1107
549 .....ACGGCTCAATACACAGACANCTCG 526
1107 roProCysGluThrHisGluThrGlyThrThr..AsnThrAlaThrThrAl 1123
525 TCTGAATCTGCGCGCT..... 507
1123 aMetSerSerValGlyAlaAsnHisGlnArgAspAlaArgAlaCysA 1140
506 .....TCTTGATCAACACACACGCGG 486
1140 laAlaGlyThrProAlaValIleArgIleSerValAlaThrGlyAlaLeu 1156
485 TCTGCCGAGAGCGA.....TT 469
1157 GlnAlaAlaGlnGlySerLysSerGlnCysGlnThrArgGlnThrSerAl 1173
468 GGTGTCCATTCGATTCGAAGATGGCAACGGCTCGCATCAGCGCAG 419
1173 aThrSerThrThrMetThrValMetAlaThrGlyAlaProCysSerAla 1190
418 GGAATTTGCTG.....AACGACGAGNTACGC 393
1190 LysProLeuLeuGlyProSerMetAlaArgLupProGlyGlyArg..... 1204
392 AGCGAGTCCCAACACCGGATGATGATGATTCGNNCANNATTCNTGCC 343
1205 .....SerProAlaPheValGlnLeuAlaProLeuSerSerLys 1217
342 GCTTAAGTTTGCCCAACGCTTCGGCGCGTGTGCTTGAACTGATTCGT 293
1217 sValArgLeuSer.....SerProSerIleLysA 1227
292 CGTTGCTT.....TCAACGCAATCAGACGCGATGAAGTAG 255
1227 spleuProAlaGlyArgHisSerHisAlaValSerThrAlaAlaMetThr 1243
254 CGCTTTTCG.....CCGCAATGGATGCGCGGATTTTCC 220
1244 ArgSerSerValGlyAlaGlyLupProArg...MetAlaProValCysG 1259
219 T.....GAACGCGCGGTAAACGACGCGCGGATCTTTGTCTT 176
1259 userLeuGlnGlyLysProSerThrThr.....ValThr 1272
175 CAACACGACTGTGCGCTTTTTCAGCGCATCGCTTCCTGACTTCATC 126
1272 alThrAlaLeuGlnAlaLeuLeuCysProSerAlaThrValThrGlnVal 1288
125 NAGGGCGCATACCGCATATTCCTTCGCCAGCAACGCGACTCGGTAAT 76
1289 CysSerAsnProProCysGluThrHisGluThrGlyThrThrAsnThrAl 1305
75 GAGC.....GGCCGCTCAATGAATGACTGTGCTGCTGCC 39
1305 aThrThrSerAsnAlaGlySerAlaGlnArgValCysSerAsnProPro 1321
seq_name: /cgn2_6/plodata/1/iaa/PCTUS_COMB.pep:PCT-US93-11721-5
seq_documentation_block:
: Sequence 5, Application PC/TUS9311721
: GENERAL INFORMATION:
: APPLICANT: Lamarco, Kelly
: APPLICANT: Wilson, Angus
: APPLICANT: Heir, Winship
: TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
: TITLE OF INVENTION: HOST CELL FACTOR
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HONBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/11721
: FILING DATE: 03-DEC-1993
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:

```



```

392 AGCGAGCTCCAAACAAACGGATTTGGATGCAGATTGCNNCANNATTCTCTGCC 343
1205 ..... SerProAlaPheValGlnIleuAlaProLeuSerIleY 1217
342 GCTTAAGTTGGCCAAACGCTTCGGGGCGGAGCTTCGAACTCGATTGGT 293
1217 sValAlaArgLeuSer.....SerProSerIleYsA 1227
292 CGTTGGCT.....TCAACGGCAATCAGACCGCACTGAAGTACG 255
1227 splenProIaGlyAlaArgHisSerHisAlaValaSerThrAlaAlaMetThr 1243
254 CGCTTTTCG.....CCGGCAGATGGATGGCGGCAATTTGCC 220
1244 ArgSerSerValGlyAlaGlyIleuProArg...MetaIaProValCysG 1259
219 T.....GAACNNGCGCGGTAACACCAACCGCCCGGATCTTTTGGCTT 176
1259 uSerIeuGlnGlyIleuSerProSerThrThr.....ValThrY 1272
175 CAACACGACACTTGGCTTTTTCAGCGGCATCGCTTCCTGACTTTCATC 126
1272 alThrAlaLeuGlnAlaLeuIleuLeuGlyProSerAlaThrValThrGlnVal 1288
125 NAGGGGCGCATACCGGCATATTCTTCGCCAAGCAACGCGACTTCGGTAAT 76
1289 CysSerAsnProProCysGlnThrHisGlnThrGlyThrThrAsnThrAl 1305
75 GAGC.....GGCCCGTCATTAAGCATCGCTCGCGGCTGGCC 39
1305 alThrThrSerAsnAlaGlySerAlaGlnArgValCysSerAsnProPro 1321
seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-270-076A-11
seq-documentation_block:
/ Sequence 11, Application US/08270076A
/ Patent No. 5667986
/ GENERAL INFORMATION:
/ APPLICANT: Sleep, Darrell
/ APPLICANT: Goodey, Andrew R
/ APPLICANT: Vakeria, Diana
/ TITLE OF INVENTION: Yeast Promoter
/ NUMBER OF SEQUENCES: 17
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: The BOC Group, Inc.
/ STREET: 100 Mountain Avenue, Murray Hill
/ CITY: New Providence
/ STATE: New Jersey
/ COUNTRY: USA
/ ZIP: 07974
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/270,076A
/ FILING DATE: 01-JUL-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: GB 8923521.2
/ FILING DATE: 18-OCT-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/597,687
/ FILING DATE: 16-OCT-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/925,286
/ FILING DATE: 04-AUG-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: SMOED, R. Hain
/ REGISTRATION NUMBER: 24864
/ REFERENCE/DOCKET NUMBER: 92HB34-3

```

```

? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 908/771-6292
? TELEFAX: 908/771-6159
? INFORMATION FOR SEQ ID NO: 11:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 806 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-08-270-076A-11

Alignment_scores:
Quality: 99.00 Length: 407
Ratio: 0.556 Gaps: 17
Percent Similarity: 43.735 Percent Identity: 22.113

Alignment block:
US-09-303-518D-127/rev x US-08-270-076A-11 ..

Align seq 1/1 to: US-08-270-076A-11 from: 1 to: 806

1322 TCACAGCACTTACCAGCAGCGGGCCATTATTCGATTATTCGCCGGGCAGAC 1273TGCGACTGA
|||||.....:||||||| |||
33 SerSeraniIeMrSerSerglyProSerSer.....Th 44

1272 GAAGCTGACAAACCGAGGTCTCT.....T 1247
||: :||| :|||
44 rPPhSerSerSerAlaThrGluserPheserThrgluThrThrValThrnp 61

1246 CGTCGCACTTCCAGCAACCAACGCT...TGCAGCTGTCGSTATCGCCG 12000
|||||.....| | | | | | | |
61 roserSerSerlyTyrrProglySerLythrGluthrSerValSerSer 77
|||||.....| | | | | | | |

1199 ACGATTAAATCGCGCAAAGAAGGATGACGAGATGATGATGACGGCATTC 1150
||| :||| :|||
78 ThrThrclunThrhllevalProthrThrThrThrThrSerValIleTh 94
|||||.....| | | | | | | |

1149 CGCGTCGTAAATGACAATGCGCACATGCGCGCGTCCACCGTTGACGG 11000
||| :||| :||| :|||
94 rProSerThrThrIleThrThrThrValcysSerThrgluThrAns 111
|||||.....| | | | | | | |

1099 CTGTCGTGAATCGAAGAGTGTGTTTTCAGAAATGCGCAGGCTGTA 1050
:||
111 er..... 111

1049 CCGGTATGAGATATTGCGCGTCCGCGCAACCCAGCAGACAGCTC 1000
||| :||| :||| :|||
112 ..AlaglyluthrThrSerlySer..... 120
|||||.....| | | | | | | |

999 TTGTGCGCGCTCTGTGATAGCAAAATGATGATGtGAGCTCCA 950
|||||.....| | | | | | | |
121 .....PolySerIleThr..... 125

949 AATATCGTGCAGCGCTGTGTAATCGCGCGCTCAATACGAAACGGAA 900
:||: :||| :||| :||| :|||
126 ....ThrThrValProcySerThrSerProsergluThrAlaserglu 140
|||||.....| | | | | | | |

899 ATCAGCGGTTGTCGCGTCAACCAATTCGCCGCACTAATTGCGATAC 850
||| :||| :||| :||| :|||
141 SerThrThrThrSerProThrProvalThrThrValalathrTh 157
|||||.....| | | | | | | |

849 TTTGCGACCC.....AAACGTFACCAAGAGCGGTF..... 816
||: :||| :||| :||| :|||
157 rValValThrThrgluTyrySerThrSerThrlnylgelygluIlet 174
|||||.....| | | | | | | |

815 .....TTGTGACTTGA 804
|||||.....| | | | | | | |

174 hrThrThrPheValThrLysAnserProThrThrThrLeuThrThrIle 190

803 GAACACCAACCAAGAACATCGCGCGTGTTCAGACGCGCTGTGCAAA 754
||| :||| :||| :||| :|||
191 AlaProthrSerSerValThr.....ThrValThrAs 201

```

```

1 FILING DATE: 05-DEC-1996
2 CLASSIFICATION: 435
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: 60/008,311
5 FILING DATE: 07-DEC-1995
6 ATTORNEY/AGENT INFORMATION:
7 NAME: Halle, Ph.D., Lisa A.
8 REGISTRATION NUMBER: 38,347
9 TELECOMMUNICATION INFORMATION:
10 REFERENCE/DOCKET NUMBER: 09010/008001
11 TELEPHONE: 619/678-5070
12 TELEFAX: 619/678-5099
13 INFORMATION FOR SEQ ID NO: 2:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 1487 amino acids
16 TYPE: amino acid
17 TOPOLOGY: linear
18 MOLECULE TYPE: protein
19 FRAGMENT TYPE: internal
20
21 US-08-760-489-2
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
100
```

```

486 .....:|||||:
      lysargproargthglyleuglnaspraplsylvalarg 500
651 .....:|||||:
      lnglylsylvalarglypheproleuargglyargglyargaspvalarg 516
500 lnglylsylvalarglypheproleuargglyargglyargaspvalarg 516
666 CACGACATTCATTCATTGAGCGCGGTGCGTCAACAAACCGTTTGA 715
      :|||:
517 lnglnhisarg.....Argcysleuserleargglyleugl 528
716 CCATCAATTATCAAGATGTATTCG.....CATC 744
      :|||:
528 nagprolysylargglysercysarglyargglyleuargproglina 545
745 GCAGCTTTGTTTGCACAGGCGG.....TCGACACGCGAGCGCGT 785
      :|||:
545 rgalaleuleuargargasnproglnglylnlaaglnhisargvalarg 561
786 GATTGCTTTGGGTGCTTCAGTCAACGACAAACACGCGCTTGGTACG 835
      :|||:
562 argasprleuarglnserlearglnlygluargleuargprarg 578
836 T.....TTGGGTGCGAA..... 848
578 glyleuarglnargserarglnvalargglyarglnvalproglin 595
849 .....AGATCGCAATTTACTGCGCGCAATTTGGTACGACAGCA 890
595 1sthrproasprleuargglysercylnglyleuargglyleuarg 611
891 CCGCGGATTTCCGTTGCGTATGAAAGCGCGATTCACACAGCGCGC 940
      :|||:
612 Alargglyphe.....Pheargargglysalarglysargh 625
941 AGCATTTATTTGGACGCGCTACCAACATCAGATTTCCGTTATGAGAAAGC 990
625 sarglyleu...glyasprleuserglylyasprleuargleuaspysarg 641
991 CGACGCAAGACGCTGTTGCGTGGTGTG.....CGCGCGACGC 1028
      :|||:
641 lyglnlnlearg..glyproglyleuargh1sarglyargasnsera 657
1029 GCACAAATTCATTCATCAGCGCTACGACCGCTCG...CCATTCCTGAAA 1075
      :|||:
657 rghpeshn1sproala1valleu..proarglnuprocyssergluas 673
1076 ACAAACTCTCAAGTTTCAGACAGCGCTCAACGCGTGGCGCGCGCATG 1125
      :|||:
673 pelyglyvalarg..glyglyleuargargglynglyleuargproleuarg 689
1126 GTGCGGATTTGACTTACGAGCGCGTAAATCGCTAGACATCTGCCTAC 1175
      :|||:
689 lylalaasprargglyleuargvalglyproargrghn1asprglyvalarg 705
1176 CCGTGTGTTGGCGATTTAATCGTGGCGGATACCGACGCGCGCAAGCAT 1225
      :|||:
706 proleuserglyserh1asnnglnly.....gluasnhrh1agllygl 720
1226 TGGTGTGTTGGAAATTCAGACAGCAAGACCGCTTGTGCACACTCGTC 1275
      :|||:
720 ylyarglylsylleuargh1sargglynglylnlaargserglyngly 737
1276 TGCCCGGCGCAATTCGA 1292
      :|||:
737 snproglyluvalarg 742
seq_name: /cgn2_6/ptodata/1/1aa_5B_COMB.pep:US-08-760-489-4
seq_documentation_block:
; Sequence 4, Application US/08760489
; Patent No. 5830696
; GENERAL INFORMATION:

```

```

? APPLICANT: Short, Jay M.
? TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
? TITLE OF INVENTION: ENZYMES
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson, P.C.
? STREET: 4225 Executive Square, Suite 1400
? CITY: La Jolla
? STATE: CA
? COUNTRY: US
? ZIP: 92037
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows95
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/760,489
? FILING DATE: 05-DEC-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 60/008,311
? FILING DATE: 07-DEC-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Haile, Ph.D., Lisa A.
? REGISTRATION NUMBER: 38,347
? REFERENCE/DOCKET NUMBER: 09010/008001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 619/678-5070
? TELEFAX: 619/678-5099
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1487 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? FRAGMENT TYPE: internal
US-08-760-489-4

```

```

alignment_scores:
      Quality: 97.50      Length: 440
      Ratio: 0.516      Gaps: 23
Percent Similarity: 42.955      Percent Identity: 23.409

```

```

alignment_block:
US-09-303-518d-127 x US-08-760-489-4 ..

```

```

Align seg 1/1 to: US-08-760-489-4 from: 1 to: 1487

```

```

219 AGGCAAAATCGCGGCAT..... 236
      :|||:
355 Argglinsnargargh1salaargalargalargaspserglu 371
237 .....CCATCGCGCGCAAAAGCGCGTACTTCACTCGTCGTC 270
      :|||:
371 rglyasprleuargproargrgh1sargalargproglnglyalagl 388
271 GTGATTCGCTGAGGCAAGCGCAAAATCGAGTTTGAACGCTACGCGC 320
      :|||:
388 euglngh1sargglyuproglproleuargh1salaargproleu 404
321 CGAAGC.....GTGG 331
      :|||:
405 Argserasprasnarglynglynglynglyproh1sargalargsp 421
332 CAACCTTAAGCGCGCGANGAANTNNGNCATGTGATTCGCGTTG 381
      :|||:
421 yargalagllygluargglylyvalargln.....Valarg.... 433
382 TGAGTGCCTGCGTANCGCGTCCGTTTCAGCAAAATCCCTGCGTGATC 431
      :|||:
434 .....glyvalhisargglyuargthrargglypro 443

```

—

355 ArgGlnAsnArgHisAlaArgAlaArgArgSerSerGlnSer 371
237CCATGCGCGCGAAAGCGCGTACTTCCGTC 270
371 rGlyAspSerLeuLeuProProArgTyrArgAlaProGlnGlyAlaGly 388
271 GGGATGGCGGTTGAAGGCAAGCAATGCGTTCGACGCGCGCGC 320
388 euGlnGlyHisArgGlnProGlnProLeuHisAlaProProLeuAlaSer 404
321 CGAAGC.....GTTGG 331
405 ArgSerAspAsnArgGlnGlyGlyProHisGlnArgAspTyrLeuGln 421
332 CAACCTTAAGCGCGGANGAANTNNGNCAATCGATCCATCCGCGTTG 381
421 yArgAlaGlyGlnArgGlyGlyAlaArgGln.....ValArg.... 433
382 TGGACTGCGCTGCGTANCCGTCGTCACAAATCCCTCGCGCGATGC 431
434GlyValHisArgGlnArgTyrArgGlyPro 443
432 CGAGCGGTCGCGATCTGTCATCGATGCGACACCAATCCGCTNCGG 481
444 ArgTyrValGlnHisLeuGlnArgAlaAspGlyArg..... 455
482 CAGACCGCTGCTGTGATCAAGAAGCGCGANGATTCAGACGANGT 531
456CysGlyAla...ArgLeuProArgAlaLeuLeuArgLeuSer 469
532 NNGCGGATTTGAGCGG...TTTGACCGAGCGTAAATCCATGCTGTGTA 578
469 IAGlyGlyTyrGlnProArgGlyGlyGlyAlaGlyAsnPro..... 482
579 GGCACCTGCGCAGACGTCGCGTCTGAATGTCGCCAACATCGAAACAC 628
483GlnHisAsp..... 485
629 ATGATTCGCGCGCGCCATCC..... 650
486LysArgProArgTyrGlnGlnAspAspGlyValArg 500
651GCCCGGTTGAGTGG 665
500 lnglyGlyArgGlyPheProLeuArgGlyArgGlyArgAspAsnLeu 516
666 CAGGACATTCATTCATTCAGCGCGTGTGCAACCAACCGTTGGA 715
517 GlnGlnHisArg.....ArgCysLeuSerLeuArgLeuGln 528
716 CCATCAATTATCAGATGTAATGC.....CANC 744
528 nArgProLysGlyArgGlnSerCysArgGlyArgGlnLeuLeuProGln 545
745 GAGCGTTGTTGCAACAGCGCG.....TCTGAACCGCGAGCGGT 785
545 rGAlaLeuLeuArgArgAsnProGlnGlnAlaGlnHisArgValArg 561
786 GATTGCTTTGGTGTCTCAAGTCAACACACCGCTTGCCTGACCG 835
562 ArgAspLeuArgGlnSerSerAlaSerGlnGlyGlnArgLeuAspArg 578
836 T.....TTTGGTGCAGAA..... 848
578 gLeuLeuHisGlnArgSerArgGlnValPheGlyAlaGlnValProGln 595
849AGTATCGCAAAATTACTGCGCGCAATGCTTGAACGACAGCAA 890
595 IsthProAspLeuLeuProGlySerSerGlnLeuArgLeuArgGln 611
891 CCGGAGATTCGCGTTGGTATTTGAACGCGCGATTTACACAGCGCGC 940
612 AlaArgGlnPhe.....PheArgArgArgGlyAlaArgGlyArgGln 625

941 ACGATTATTTGGAGCGCTTACCAATCAAGATTTCCGTTATCGAAGAGC 990
625 sArgLeu...GlyAspLeuSerGlyGlyAspLeuArgLeuAspArgG 641
991 CGCAGCAAGAGCTGTGCGTGGTGG.....CGCCGACGCC 1028
641 lylGlnGlnLeuArg...GlyProGlyLeuArgHisArgGlyAspAsnSer 657
1029 GGACAAATATCCATCAACGCGCTACGACCGTGG...CCATTTCTGAATA 1075
657 rGPhAsnHisProAlaAlaValLeu...ProArgGlnProCysSerGln 673
1076 ACAAACTTCAAGTTCACGACAGCGCTCAAGCGTGGCGACCGCGCATG 1125
673 pGlyGlyValArg...GlyGlyLeuArgArgGlnGlnGlyLeuProLeuLeuG 689
1126 GTGCCGATTTGTAATTCAGCGCGCTAATGCGCTAGACATTCGCTTAC 1175
689 lylAlaAspArgGlnLeuArgValGlyProArgPheGlnAspGlyValArg 705
1176 CCTGCTTTTCGCGATTTAATGTCGCGATACGACAGCGCGAGCAT 1225
706 ProLeuSerGlySerHisAsnGlnGly.....GlnAsnTrpAlaGlyGln 720
1226 TGGGTTGCTTGAATTCGACGAAGAAGACCTGCTTGTGCGAGCTTGTGTC 1275
720 yLysArgGlyGlyLeuGlnHisArgGlyGlnGlnArgSerGlnGlnGly 737
1276 TGGCGCGCAATTCAGA 1292
737 snProGlyGlyValArg 742

seq.name: /cgn2_6/prodata/1/iaa/6B_COMB.pep:US-09-185-373-4

seq_documentation block:
; Sequence 4: Application US/09185373
; Patent No. 635179
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185,373
; FILING DATE: 03-NO-6335179-1998
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,489
; FILING DATE: 05-DEC-1996
; APPLICATION NUMBER: 60/008,311
; FILING DATE: 07-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Halleg, Ph.D. Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:


```

;
; LENGTH: 1487 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; FRAGMENT TYPE: internal
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-185-373-4

```

```

alignment_scores:
  Quality: 97.50      Length: 440
  Ratio: 0.516       Gaps: 23
  Percent Similarity: 42.955   Percent Identity: 23.409

```

alignment_block:

US-09-303-518d-127 x US-09-185-373-4

Align seg 1/1 to: US-09-185-373-4 from: 1 to: 1487

```

219 AGGCAAAATCGCCGCAT..... 236
    |||||
355 ArgGlnAsnArgArgHisAlaArgAlaArgAspSerGluSer 371
    |||||
237 ..... 270
    |||||
371 rGlyAspSerLeuLeuProArgTyrArgLarProGlnGlyAlaGly 388
    |||||
271 GTGATTCGCGTGAAGCAAGCAAGATTCGATTCGATTCGATTCG 320
    |||||
388 euGlnGlnHisArgGluProGlnProLeuHisAlaProProLeuLaser 404
    |||||
321 CGAAGC..... 331
    |||||
405 ArgSerAspAsnArgGluGluGluGluProHisGlnArgAspTyrLeu 421
    |||||
332 CAACTTAAGCGCGGANGAANTNGNNGCATTCGATTCGATTCGATTCG 381
    |||||
421 YArgLarGluGluArgGluGluValArgGln.....ValArg.... 433
    |||||
382 TGAGTGGCGTGGTTCGATTCGATTCGATTCGATTCGATTCGATTCG 431
    |||||
434 .....GlyValHisArgGluGluGluGluGluGluGluGluGluGlu 443
    |||||
432 CGAGCGTTCGCGCATTCGATTCGATTCGATTCGATTCGATTCGATTCG 481
    |||||
444 ArgGlyValGlnHisLeuGlnArgAlaAspArgGlu..... 455
    |||||
482 CAGAGCGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 531
    |||||
456 .....CysGlyAla.....ArgLeuProArgAlaLeuLeuArgLeuSer 469
    |||||
532 NTGCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 578
    |||||
469 IarGlyGluGluProArgGluGluGluGluGluGluGluGluGluGlu 482
    |||||
579 GCGAGCTGGCGCGAGCGATTCGATTCGATTCGATTCGATTCGATTCG 628
    |||||
483 .....GlnHisAsp..... 485
    |||||
629 ATGATTCGCGCGCGCGCATTC..... 650
    |||||
486 .....LysArgProArgThrGluGlnAspArgGluGluGluGluGlu 500
    |||||
651 .....GCGCGTTCGATTCGATTCGATTCGATTCGATTCGATTCG 665
    |||||
500 InGlyGluGluArgGluPheProLeuArgGluGluGluGluGluGluGlu 516
    |||||
666 CAGCGCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 715
    |||||
517 GlnGlnHisArg.....ArgCysLeuSerLeuArgLeu 528
    |||||
716 CCATCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 744
    |||||

```

```

528 narProLysGluArgGluSerCysArgGluGlnLeuProGln 545
745 GGAGCTTTGTTTGCACAGCGCG.....TGTGACACCGAGCGCT 785
    |||||
545 rGlaLeuLeuArgArgAsnProGlnGlnAlaGlnHisArgValArg 561
    |||||
786 GATTCGTTGGTGGTTCGATTCGATTCGATTCGATTCGATTCGATTCG 835
    |||||
562 ArgAspLeuArgGlnSerSerAlaSerGlnGluGluGluGluGluGlu 578
    |||||
836 T.....TTTGGTGGCA..... 848
    |||||
578 GLeuLeuHisGluArgSerArgGlnValPheGlyAlaGlnValProGlu 595
    |||||
849 .....AGATCGCAATTCATTCGCGCGGCAATTCGATTCGATTCG 890
    |||||
595 IsthProAspLeuProGlySerSerGlnLeuArgLeuGln 611
    |||||
891 CCGGCTGATTCGCGTTCGATTCGATTCGATTCGATTCGATTCGATTCG 940
    |||||
612 AlaArgGluPhe.....PheArgArgArgGlyAlaArgGlyArgGln 625
    |||||
941 ACGATTAATTCGAGCGTTCACACATTCGATTCGATTCGATTCGATTCG 990
    |||||
625 sarGlu.....GlyAspLeuSerGluGluGluGluGluGluGluGlu 641
    |||||
991 CGCAGCAAGAGCTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCG 1028
    |||||
1029 GGCACAAATATTCATTCACGCGTTCGAGCGCGTTCGATTCGATTCG 1075
    |||||
641 GlnGlnGlnLeuGluGluGluGluGluGluGluGluGluGluGluGlu 657
    |||||
657 rGlnHisArgProAlaValLeuProArgGluProCysSerGln 673
    |||||
1076 ACAAACTTCAGTTCACAGCGCGTTCACAGCGCGTTCACAGCGCGCG 1125
    |||||
673 rGlyGluValArg.....GluGluLeuArgArgGlnGluGluGluGlu 689
    |||||
1126 GTCGCGATTCGATTCATTCAGCGCGGATTCGCGGATTCGATTCGATTC 1175
    |||||
689 YAlaAspArgGlnLeuArgGluGluGluGluGluGluGluGluGluGlu 705
    |||||
1176 CCGCTTCGCGGATTCATTCGCGGATTCGCGGATTCGCGGATTCGCGG 1225
    |||||
706 ProLeuSerGlnSerHisGlnGlu.....GlnsntHrAlaGln 720
    |||||
1226 TGGGTTGCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1275
    |||||
720 YLysArgGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 737
    |||||
1276 TGCCTGGCGCAATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1292
    |||||
737 snProGlyGluValArg 742

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.dep:US-09-081-345-18

seq_documentation_block:

```

; Sequence 18, Application US/09081345
; Patent No. 6228641
;
; GENERAL INFORMATION:
; APPLICANT: Bahija Jallal
; APPLICANT: Gregory D. Florman
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
; TITLE OF INVENTION: PTPO4 RELATED DISORDERS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066

```

COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 MEDIUM TYPE: storage
 COMPUTER: IBM compatible
 OPERATING SYSTEM: IBM P.C. DOS 5.0
 SOFTWARE: Fastseq for Windows 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/081.345
 FILING DATE: Herewith
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/047,222
 FILING DATE: May 20, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Warburg, Richard J.
 REGISTRATION NUMBER: 32,327
 REFERENCE/DOCKET NUMBER: 234/253
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (213) 489-1600
 TELEFAX: (213) 955-0440
 TELEX: 67-3510
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 802 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-09-081-345-18

alignment_scores:
 Quality: 95.00 Length: 306
 Ratio: 0.638 Gaps: 14
 Percent Similarity: 48.693 Percent Identity: 21.895

alignment block:

US-09-303-518D-127/rev x US-09-081-345-18

Align seg 1/1 to: US-09-081-345-18 from: 1 to: 802

```

810 GACTTGAGAACCAACCAACCAAT..... 787
335 AspValIysThrIAsnGlnHisSerLysGlnGlyAlaGluAlaGluSe 351
786 .....CAGCGCGTGGTGTTCAGACGGCGCTGTGCAACCAACG 747
351 rThGlyGlySerSerLeuGlyLeuAlaThrSer.....Thm 364
746 CCATGGCAATTACATCTGTGATTAATGATGTCACAAAC.....G 706
364 etAsnAlaGlnGluGluValLeuHisSerAlaLysSerSerProSer 380
705 TTGTGTTGACCGACCGCGCTCAATGAATGAATGCGTGGCC.....AC 662
381 PheAsnGlyLeuGluLeuAsnGlyCysAsnAsnLysAlaValIleTh 357
661 TCAACCGCGCGCGATGCGCGCGCC.....G 636
397 rArgSngGlnAlaIAsnGlnAlaSerProValValGlyLProLeuGln 414
635 AATTCATGCTTGGATGTTGGACGATTTTCAGACGGCGACGTCGCGC 586
414 yStyGlnSerLeuAspPheGlySerMetLeu.PheGly...SerCysPr 429
585 AGCTGCTTACACACATGATGATTTAAGCTGGTCAACAGGCTCAATACA 536
429 oSerAlaLeuPro.....IleAsnThrAlaAspArgTyrHisAsnS 443
535 GC.....ANACNTGCTGTGAATCAGTGCAGCGGCTCTTGTGATC... 498
443 eLysGlyProValLysArgThrLysSerThrProPheLeuIleGln 459

```

```

497 ...ACAACCAAGGCTGTGCCGCAACGCA..... 471
460 GlnAlaGlyThrAsnAspLeuAlaValGlyAspLysPheSerCysLeuG 476
470 .....TTGGTGTCCATCGATTGACGAGA 446
476 uSerGlnLeuHisGlnHisTyrSerLeuAlaGlyLeuGlnValGlnArg 493
445 TGGCGAACGGCTCGCATCGACGCGAGGATTTTGGTGAACGCGAGNA 396
493 aLAlaHisValSerSerGlnGluLeuAsnTyrSerLeuProGlyAlaCys 509
395 CGCAGCGCAGTC.....CACAAACGGATTTGATGATGATTCGANN 355
510 AsnAlaSerCysValProAlaGlnHisSerProGlyAlaLeuAlaHis 526
354 NANTTCNTGCGCGCTTAAGTTTCCCAACGCTTCGGCGCGGTAGCTTGA 305
526 uTyrThrSerLeuAlaGlnAspProTyrPheSer..... 537
304 ACTGATTTGCTGTGCTTCACGCGCATATCAGACGAGTGAAGTACG 255
538 .....SerSerProAsnSerAlaAspSerLys..... 547
254 CGCTTTTCGCGCGGATGATGCGCGCGATTTTGCCTGAAACGCGCGGT 205
548 .....MetSerPheAspLeuProGlnLysGlnAsp 558
204 AACACCAAGCGCGGATNTTTTGTCTTCAACAGCATTTGGCTTTT 155
558 yAlaThrSerProGlyAlaLeuProAlaSerSerThrThrSerPhe 575
154 TGACGCGATCGCTTCTTGAATTCATCNAAGGGCGCATACCGGCATAT 105
575 heTyrSerAsnProHisAspSerLeuValMetAsnThrLeuThrSerPhe 591
104 TCTTGCAGACGCGGACACTTCGATATAGCGGCCGCTCATTAATATGAC 55
592 SerProProLeuAsnGlnGluThrAlaValGlnAlaProSerAlaArgTh 608
54 TTGCTCGGCTTGCC 39
608 rAspAspGlnIlePro 613

```

seq_name: /cgn2_6/plodata/1/iaa/6A.COMB.pep:US-09-335-409-5

seq_documentation_block:

```

; Sequence 5, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-335-409-5

```

alignment_scores:
 Quality: 93.50 Length: 538
 Ratio: 0.408 Gaps: 29
 Percent Similarity: 42.565 Percent Identity: 21.190

alignment_block:
 US-09-303-518D-127 x US-09-335-409-5 ..

Align seg 1/1 to: US-09-335-409-5 from: 1 to: 7257

```

61 TATGACGGGCGCCGATACCGAAGTCGCGTGTGGCGAA...GATA 107
   ||||| ||| ||||| ||||| ||||| ||||| |||||
6356 PHEASPGYGLNALAProthSerValHisLeuGlySerLeuAspGI 6372
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
108 TCCCGGTATGCGCCCTTGATGAAGTCAAGAGCGCATCCCTCAAA 157
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6372 uArgGlyValLeuAspAlaAspAlaProPheAspAlaLeuGIug 6389
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
158 AAGGC.....CAAGTGTGTTGAAGACAAAGNAT 189
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6389 lSerLeuValArgGlyCysAspSerValLeuTrpThrValGlnAlaVal 6405
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
190 CCGGGCGGTGTATTACCGGCCN.....GTTTCAGCGCA 224
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6406 AlAGLYAlAGLYPheArgAspProArgLeuTrpLeuValThrArgGI 6422
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 AATCGCGGCATCCATCGCGGAAAGCGCGTACTTCAGTCGCTC... 270
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6422 yAlAGlnAlAlIeGLYAlAGLYAspValSerValAlAGlnAlaProLeuL 6439
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
271 .....GTGATTGCCGTGAA..... 285
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6439 euGIyLeuGIyArgValIleAlaLeuGIuHisAlAGLYLeuArgCysAla 6455
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
286 .....GGCAAC...GAGCAATGCA 302
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6456 ArgIleAspLeuAspProAlaArgArgAspGIyGluValAspGIuLeuLe 6472
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
303 GTTGCAGCGGTACGCGCGCGAAGCG.....TTGCCAACTTAAGCG 343
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6472 uAlAGLYLeuLeuAlaAspAspAlaGIuGIuValAlaPheArgGIyG 6489
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
344 GCGANGAANTNGNNGCAATCTGATCCATCCGTTTGGACTGGCGTG 393
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6489 lYGIuArgArgValAlaArgLeuValArgArgLeuProGIuThrAspCys 6505
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
394 CGTANCGTCCGCTTCAGCAAAATCCCGCGTATGCCGCGCGTCCG 443
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6506 Arg.....GIuLysIleGIuProAlAGIuGIyArgProPheArg 6518
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
444 CATCTGCTCAAT.....GCCA 460
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6518 glEuGIuIleAspGIySerGIyValLeuAspLeuValLeuArgAlaIAT 6535
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
461 TGGACCAATCCGCTNCGCGAGACCCGTGTGTGATCAAAAGAGCG 510
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6535 hngIuArgArgProProGIyProGIyGIuValGIuIleAlaValAla 6551
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 GNC.....GANGATTTCAGACGANGTCTGTTGATTCAGCGC 548
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6552 AlAGLYLeuAsnPheLeuAspValMetArgAlaMetGIyLeuTrpGI 6568
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 TTTGACGAGCGGTAAATTCATGTG.....TGTAAAG..... 579
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6568 yProGIyAspGIyProValAlaLeuGIyAlAGIyCysSerGIyArgIleV 6585
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
580 ..GCAGCTGGCGGACAGCTGCCGTCT.....GAAATGCTGCC 615
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6585 aAlaMetGIyGIuGIyAlAGIuSerLeuArgIleGIyGIuAspValAl 6601
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
616 AACATCGAAACATGAATTCGCG..... 639
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6602 AlAValAlaProPheSerPheGIyThrHisValThrIleAspAlaArgMe 6618
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 ....GGCCGCGATCCGGCGGTTTAGAGGCGACGACATTCATTCAATG 685
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6618 tLeuAlaProArgProAlaAlaLeuThrAlaAlAGlnAlaAlaLeu. 6634

```

```

686 AGCCGGTCGGTGCAGAAACAAACCGTTTGACCATTCATTCAGATGTA 735
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6635 ..ProVal...AlaPheMetThrAlaTrp.....TYRGIyLeu 6645
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
736 ATTGCCATCGAGCGTTTGTTCACACAGCGCGTCTGAAACCGACCGCT 785
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6646 ValHisLeuGIyArgLeuArgAla.....GIyGIuArgyA 6657
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
786 GATT.....GCTTGGGTGTTCTCAAGTCAACAAACCGCGCTCT 826
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6657 lLeuIleHisSerAlaThrGIyGIyThrGIyLeuAlaAlaValGIleAl 6674
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
827 TCCGTCACGGTTTGGTGGTCAAGTA..... 852
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6674 lArgHis...LeuGIyAlaGIuIlePheAlaThrAlAGIyThrProGIu 6689
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
852 ..... 852
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6690 lYsArgAlaTrpLeuArgGIuGIuGIyIleAlaHisValMetAspSerArg 6706
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
853 .....TGCCAATT.....ACTGCGGCGCAATTGG 877
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6706 gSerLeuAspPheAlaGIuGIuValLeuAlaAlaThrIySGIyGIuGIy 6723
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
878 TTGACGCGACAGCAACCGCGTATTCGCGTTCGGATTCAGACGGCGCAT 927
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6723 AlAspValValLeuAsnSerLeuSerGIyAlaAlaIleAspAlaSerLeu 6739
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
928 ACACAGAGCGCGCAGCATTTATTGGACCGTCCACATTCAGATTCCGT 977
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6740 Ser.....ThIle 6742
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
978 TATCGAAGAGCGCGCAGCAAGAGCTGTTCGCTGGTGGCGCGCAGC 1027
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6742 uValProAspGIyArgPheIleGIuLeu.....GIyLysT 6754
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
1028 CGGACAAATATCTCATCAGCGGTACGACACCTCGGCATTCCTGAANAAC 1077
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6754 hAspIleIyArgAlaAspArgSerLeuGIyLeuAlaHisPheArgLysSer 6770
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
1078 AAATCTTCAGATTCACGACGACCGGTCAACGCGGCGACCGCGCATGCT 1127
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6771 LeuSerIy.....SerAlaVal.....AspLeuAlAGIyLe 6781
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
1128 GCCGATTGGTACTTACGAGCGCGTA..... 1152
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6781 uAlaValArgArgProGIuArgValAlaAlaLeuLeuAlAGIuValAla 6798
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
1153 .....ATGCCGTAGACATCCGTGCT 1173
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6798 sPLeuLeuAlaArgGIyAlaLeuGIuProLeuProValGIuIlePhePro 6814
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
1174 ACCCTGCTTTGGCGGATTAATCGTCGCGGATACCGGACGCGCGAC 1223
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6815 LeuSerArgAlaAlaAspAlaPhe..ArgLysMetAlAGlnAlaGlnHis 6830
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
1224 ATTGGGTTGCTTGAATTCGACGAGAAAGACCTGCTTTGCAAGCTTGC 1273
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6831 LeuGIyLysLeuValIleuAlaLeuGIuAspProAsp...ValArgIleArg 6846
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
1274 TCTGCGCGCGGCA 1286
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
6846 gValProGIyGIu 6850
   ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: /cgn2_6/putdata/1/laa/5B_COMB.pep:US-09-568-102-5

seq_documentation_block:

```

; Sequence 5, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James

```



```

6723 alaspvalleuasnserleuserglyalalaleasplaserleu 6739
928 ACACAAGGCGCGCAGATTATTTGGACGCTACACAAATCAGATTTCCT 977
6740 Ser.....Thre 6742
978 TATCGAAGAGCGCGCAGCAAGAGCTGTTCGGCTGGTGGCGCCGAC 1027
6742 uvalproaspglyargphei1le1leu.....gly1st 6754
1028 CGGCAAAATATCTCATCCAGCGGTACGACCCCTCGCCATTTCGAAAAAC 1077
6754 hraspllelyralaasparserleuglyleu1ahisphearglysser 6770
1078 AAATCTTCAAGTTACGACAGCGCTCAACGGTGGCGACCGCCCATGT 1127
6771 leuser Tyr.....ser1aVal.....Aspleu1alegly1e 6781
1128 GCGGATTGGTACTTACGACGCGTA..... 1152
6781 ualavalargargprogluargvalalalaleu1aleu1agluvala 6798
1153 .....ATGCGCTAGACATCCGCT 1173
6798 spleu1eualargglyalaleu1euglnproleuprovale1uilephep 6814
1174 ACCCTCTTTGGCGGATTAAATCGCGCATACCGACAGCGCCCAAGC 1223
6815 leuserargalalalasp1alphe..Arglymet1alagl1alagl1his 6830
1224 ATTGGTGTCTGGAAATGGACGAGAAAGACCTCGTTGTGCACCTTCG 1273
6831 leuglylyleuvalleu1aleu1leu1asproasp...Valarg1leat 6846
1274 TCTGCCCGGGCA 1286
6846 gyalprogly1u 6850

```

seq_name: /cgm2_6/ptodata/1/iaa/6B_COMB.pep:US-09-568-480-5

```

seq_documentation_block:
: Sequence 5, Application US/09568480
: Patent No. 6355458
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,480
: PRIOR FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 5
: LENGTH: 7257
: TYPE: PRT
: ORGANISM: Sorangium cellulosum
US-09-568-480-5

```

```

alignment_scores:
Quality: 93.50 Length: 538
Ratio: 0.408 Gaps: 29
Percent Similarity: 42.565 Percent Identity: 21.190

```

alignment_block:
US-09-303-518D-127 x US-09-568-480-5 ..

Align seg 1/1 to: US-09-568-480-5 from: 1 to: 7257

```

61 TATAGCGGCGCCGTCATACGAGTGGCGTGTGCCAA...GATA 107
6356 Pheasp1ygl1alalaprothserValHis1leu1ys1leuasp1 6372
108 TGCCGTATGCGCCCTNGATGAAGTCAAGAAAGCGATGCCGTA 157
6372 uArg1yValleuasp1alasp1alproph1easplalaleu1eug 6389
158 AAGC.....CAAGTCTGTTTGAAGCAAAAGNAT 189
6389 luserleuValarg1yCysaspserValleut1r1h1rVal1gl1alaval 6405
190 CCGGCGCGGTGTTCACCGCGCN.....GTTCCAGGCA 224
6406 Alag1y1alag1yPheargasp1ro1ar1leut1r1eual1h1rarg1 6422
225 AATCGCGCCATCCATCGCGCGCAAAAGCGGTACTTCACTCGTC.... 270
6422 yAlag1alal1e1y1alag1yasp1alserValalag1alaleu1 6439
271 .....GTGATTCGCGTTGA..... 285
6439 euglyleu1yargval1le1alaleu1h1salag1leu1argCysAla 6455
286 .....GGCAAC...GAGCAAAATCGA 302
6456 Arg1leasp1euaspro1alalargasp1ygl1valasp1leu1e 6472
303 GTTCGAAGCTTACGCGCCGCAAGCG.....TTGGCAAACTTAAGC 343
6472 uAlag1leu1eualasp1alasp1alag1u1alalapheargly 6489
344 GCGANGAANTNNGCAATCGATCCATCCGCTTGTGCAGTGGCGTC 393
6489 lycluar1argvalAlalarg1eualarg1leuprog1u1r1aspCys 6505
394 CGTANCCGTCGTTCAAGCAAAATCCGCGCATGCCGAGCGCTTCCG 443
6506 Arg.....Gly1s1le1u1pro1alag1u1yarg1rothear 6518
444 CATCTGCTCAAT.....GCCA 460
6518 gleu1u1leasp1yserglyValleuasp1eualleu1eualarg1at 6535
461 TGGACACCAATCCGCTNCGCGGACACCGCTGTTGATCAAGAGCC 510
6535 hrg1uar1arg1ro1rog1y1rog1ygl1u1alag1u1alavalglua1a 6551
511 GNC.....GANGATTTCAAGCANGTGTGATTTAGCGG 548
6552 Alag1y1euas1nph1euas1pValmetarg1alame1y1le1y1ro1 6568
549 TTTGACCGAGCGTAAATCCATGTG.....TGTAA..... 579
6568 yProgl1yasp1yProvalalaleu1y1alag1u1y1s1er1yarg1ley 6585
580 ..GCAGTGGCGGACGATGCCGCT.....GAAATGCTGCC 615
6585 alalame1ygl1u1y1alag1u1ser1leuarg1le1y1nasp1alval 6601
616 AACATCGAAACATCATGATTCGCG..... 639
6602 AlavalalalaprotheserPhegly1ThrHisVal1Th1leasplalarghe 6618
640 ....GGCCGCGATCCGCGGTTTGAAGTGGCAAGCATTTTCATTTG 685
6618 l1eualalaproarg1ro1alalaleu1h1alag1u1alalaleu1e 6634
686 AGCCGCTGGTGCAGCAAAACCGTTTGGACCATCATATATCAAGATGTA 735
6635 ..Proval...Alaphemet1r1al1atrp.....Tyr1glyleu 6645

```

```

736 ATTGCATCGACGTTTGTTCGACACAGCCGCTTGAACACCGAGCGCT 785
    ::::::::::::::::::::
6646 VAIHSLHSLGGLYARGLEUARGLA.....GLYLUALRGLA 6657
786 GATT.....GCTTGGGTGGTTCACAGTCAACAACACCGCCCT 826
    ::::::::::::::::::::
6657 LLEUHLHSLSERIALATHRGILYTHGLYEUHALAVALAGLNLIA 6674
    ::::::::::::::::::::
827 TCGTACCGTTTGGGTGCGAAGTA..... 852
    ::::::::::::::::::::
6674 LARGHLS..LEUGLYALAGLILIEPHEALATHRALGLYTHRPROGL 6689
    ::::::::::::::::::::
852 ..... 852
6690 LYSARGALATRPLEUARGLUGLNLGLYLEALHLSVALMETASPERAR 6706
853 .....TCGCAAT.....ACTGGGCGGAATTGG 877
6706 GSERLEUASPHEALAGLUGLNLVALLEUALAATHRYSGLYGLY 6723
878 TTGACGACAGACACCGCGTATTCGGTTCGTAATTAACGCGCGATT 927
    ::::::::::::::::::::
6723 ALASPVALLEUASNSERLEUSERGLYALALALILEASPALASERLEU 6739
928 ACACAAAGCGCGACGATTATTGGACCGCTACACATCAGATTCCGT 977
    ::::::::::::::::::::
6740 SER.....THRE 6742
978 TATGACGAAGACCGCGACGAAGAGCTGTTCGGTGGTGGCGCGCAGC 1027
6742 UVALPROBSPGLYARGPHEILEGLU.....GLYLYT 6754
1028 CGGACAAATACCTCCATCAGCGCTACGACCCGCGCATTTCTGAAAAC 1077
    ::::::::::::::::::::
6754 HRASPLIERYRLAASPARSERLEUGLYLEUALHLSPHEARGLYSSER 6770
1078 AAACCTTCACGTTCAAGACAGACGCGCAACGCGGACCGCGCATGTG 1127
6771 LEUSERTYR.....SERIALVAL.....ASPLEUALGLYLE 6781
1128 GCGGATGTGACTTACGAGCGCGTA..... 1152
6781 UALVALARGARGPROGLUARGVALAALALEULEUALGLUVALA 6798
1153 .....ATGCCCTAGACATCTGCGCT 1173
6798 SPLEULEUALARGLYALALEUGLNPLOEUPROVALGLULIEPHEPRO 6814
1174 ACCCTGTTTTGGCGATTAAATCGTGGGATACGACGACGCGCAGC 1223
6815 LEUSERARGALALAASPALAPHE..ARGLYSMETALAGLINALGLNLIS 6830
1224 ATTGGTGTGCTTGAATTTGACGAAGAAGCTCGCTTTGTGACGCTTG 1273
    ::::::::::::::::::::
6831 LEUGLYLSEUVALLEUALALEUGLUASPROASP...VALARGILEAR 6846
1274 TCTGCCCGGGCAA 1286
6846 GVALPROGLYGLN 6850

```

seq_name: /cgn2_6/ptodata/1/laa/GB_COMB.pep:US-09-568-486-5

```

seq_documentation_block:
; Sequence 5, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern

```

```

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-568-486-5

```

```

alignment_scores:
  Quality: 93.50      Length: 538
  Ratio: 0.408      Gaps: 29
  Percent Similarity: 42.565      Percent Identity: 21.190

```

alignment_block:

US-09-303-518D-127 x US-09-568-486-5 ..

Align seq 1/1 to: US-09-568-486-5 from: 1 to: 7257

```

61 TATGACGCGCGCGTATTCACGAAGTCGCTTGGCGAA...GATA 107
    ::::::::::::::::::::
6356 PHEASPGLYGINALAPROTHRSERVALHLSLEUGLYSERLEUASPGI 6372
108 TGCCGATATGCGCCCTNGATGAAAGTCAGAAAGCGGATCGCTCAAAA 157
    ::::::::::::::::::::
6372 UARGGLYVALLEUASPALAASPALAPROPHASPALASPALALEUG 6389
158 AAGGC.....CAAGTGTGTGAAGACAAAGNAT 189
6389 IUSERLEUVALARGLYCYSPASPERVALLEUTRPTHVALGLNALVAL 6405
190 CGGCGCGTGTGTTTACCGCGCN.....GTTTCAGCGCAA 224
6406 ALAGLYALAGLYPHEARGSPROPRORGLREUVALTRHARGI 6422
225 AATCGCCGCATCCATCGCGCGGCAAGCGGCTACTTCAAGCGGTC... 270
6422 YALAGLINALIEGLYALAGLYASPVASERVALALAGLINALPLEUL 6439
271 .....GTGATGGCGCTTGAA..... 285
6439 EUGLYLEUGLYARGYVALLEALALEUGLNLHLSALAGLULEUARGYALA 6455
286 .....GGCAAC...GACGAATGCA 302
6456 ARGILEASPLEUASPPROALARGARGSPGLYGLUVALASPGLULEUL 6472
303 GTTCGACACCTACGCGCGCGAAGCG.....TTGCAAACTTAACG 343
6472 UALAGLULEUVALAASPSALAGLUGLNUVALAALAPHEARGLYG 6489
344 GCGANGAANTNNGCAATGTGATCAATCCGCTTGTGACGCGCTG 393
    ::::::::::::::::::::
6489 LYGLUARGARGVALAALARGLEUVALARGARGLEUPROGLUNHRASPCYS 6505
394 CGTANCCGTCGCTTACGAAATCCCTGCCCTCGATGCCGACGCGCTGCG 443
6506 ARG.....GLULYSILEGLUPROVALAGLUGLYARGPROPHEAR 6518
444 CATCTGTCGTCAT.....GCGA 460
6518 GLEUGLULIEASPGLYSERGLYVALLEUASPSPLEUVALLEUARGALAT 6535
461 TGACACACATCGCTNGCGCGACGACCGCTGCTGTGATCAAGAAGCC 510
    ::::::::::::::::::::
6535 HRCGLUARGARGPROPROGLYPROGLYGLUVALGILULEALVALAGIUALA 6551

```

```

511 GNC.....GANGATTTCAGACGANGNTGCTGATTGAGCCG 548
    |||
5552 AlaGlyLeuAspPheLeuAspValMetArgAlaMetGlyIleTyrProG1 6568
    |||
549 TTGACCGAGCGGTAATCCATGTC.....TGTAAG..... 579
    |||
6568 yProG1AspGlyProValAlaLeuGlyAlaGlyCysSerGlyArgIleV 6585
    |||
580 ..GCAGCTGGCGCAGACGTCGCGTC.....GAAATGCTGCC 615
    |||
6585 alAlaMetGlyGlyLeuGlyValGlySerLeuArgIleGlyLeuAspValV 6601
    |||
616 AACATCGAAGACATGATTCGCG..... 639
    |||
6602 AlaValAlaProPheSerPheGlyThrHisValThrIleAspAlaArgMe 6618
    |||
640 ....GGCCCGCATCCGCGCGGTTGAGTGGCAGCAGCATTCATTTCATG 685
    |||
6618 tLeuAlaProAlaGlyProAlaAlaLeuThrAlaAlaGlnAlaAlaLeu. 6634
    |||
686 AGCCGGTCGGTGCAGCAAAACCGTTTGGACCATCATATTCAGAGTGA 735
    |||
6635 ..ProVal...AlaPheMetThrAlaTrp.....TyrGlyLeu 6645
    |||
736 ATTGGCATCGAGCGTTTGTTCGACAGAGCGCGTCTGAACAGCGCGCT 785
    |||
6646 ValHisLeuGlyArgLeuAlaArgAla.....GlyIleArgVal 6657
    |||
786 GATT.....GCTTGGGGGTCTCANGTCACAAACAGCGCGCT 826
    |||
6657 lLeuIleHisSerAlaThrGlyGlyIleThrGlyLeuAlaAlaValGlnIleA 6674
    |||
827 .TGGTACCGTTTGGTGGCAAGTA..... 852
    |||
6674 lArgHis...LeuGlyAlaGlnIlePheAlaThrAlaGlyThrProGlu 6689
    |||
852 ..... 852
    |||
6690 LysArgAlaTrpLeuArgGlyGlnGlyIleAlaHisValMetAspSerAr 6706
    |||
853 .....TCGCNAATT.....ACTGGCGGCAATGG 877
    |||
6706 gSerLeuAspPheAlaGlnGlnValLeuAlaAlaThrLysGlyGlnGlyV 6723
    |||
878 TTGACGAGACAGACCGCGGTATTCGCGTTCGATTGAACGGCGCGATT 927
    |||
6723 alAspValValLeuAsnSerLeuSerGlyAlaAlaIleAspAlaSerLeu 6739
    |||
928 ACACAGCGCGCGCATTTTGGGACGCTACACACATGAGATTTCGCT 977
    |||
6740 Ser..... 977
    |||
978 TATCGAAGAGCGCGCAGCAAGAGCGTTCGCGTGGTTCGCGCGCAGC 1027
    |||
6742 uValProAspGlyArgPheIleGlnLeu.....GlyLysT 6754
    |||
1028 CGAGCAAAATCTCCATCAGCGCGTACAGACCTCGCCATTCTCGAAGAAC 1077
    |||
6754 hrAspIleTyrAlaAspArgSerLeuGlyLeuAlaHisPheArgLysSer 6770
    |||
1078 AAACCTTTCAAATTTCAGACAGCGCGTCAAGCGGTGGCGCGCATGTGT 1127
    |||
6771 LeuSerTyr.....SerAlaVal.....AspLeuAlaGlyLe 6781
    |||
1128 GCGGATTGGATTACTACAGCGCGTA..... 1152
    |||
6781 uAlaValAlaArgArgProGluArgValAlaAlaLeuLeuAlaGlnValVala 6798
    |||
1153 .....ATGCCGTAGACATCCTGCGCT 1173
    |||
6798 splLeuLeuAlaArgGlyAlaLeuGlnProLeuProValGlnIlePhePro 6814
    |||
1174 ACCCTGCTTTTGGCGGATTATTCGTGCGGATACGACAGCGCGCAGC 1223

```

```

6815 LeuSerArgAlaAlaAlaPhe..ArgIleMetAlaGlnAlaGlnHis 6830
    |||
1224 ATTGGTGTGCTGGATTGGACGAGAGACCTCGCTTGTGAGCTTCG 1273
    |||
6831 LeuGlyLysLeuValLeuAlaLeuGlnAspProAsp...ValArgIleAr 6846
    |||
1274 TCTGCCCGCGGCA 1286
    |||
6846 gValProGlyGlu 6850

```

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-09-568-472-5

```

```

seq_documentation_block:
; Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyrt, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR APPLICATION NUMBER: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5

```

```

alignment_scores:
  Quality: 93.50      Length: 538
  Ratio: 0.408      Gaps: 29
  Percent Similarity: 42.565      Percent Identity: 21.190

```

```

alignment_block:
US-09-303-518d-127 x US-09-568-472-5 ..

```

```

Align seg 1/1 to: US-09-568-472-5 from: 1 to: 7257

```

```

61 TATGAGGCGCGCGCATTAACGAGTCGCTTGGTCGCA...GAATA 107
    |||
6356 PheAspGlyGlnAlaProThrSerValValHisLeuGlySerLeuAspG1 6372
    |||
108 TGGCGGTATGCGCCCTNGATGAAGTCAGAGAGGCGATGCCGCAAA 157
    |||
6372 uArgGlyValLeuAspAlaAlaProPheAspAlaAspAlaLeuGln 6389
    |||
158 AAGC.....CAAGTCGTGTTGAAGACAAAGANAT 189
    |||
6389 lSerLeuValAlaArgGlyCysAspSerValLeuTrpThrValGlnAlaVal 6405
    |||
190 CCGGCGGTGTTTACCGCGCCN.....GTTTCAGGCAA 224
    |||
6406 AlaglyAlaGlyPheArgAspProArgLeuTrpLeuValThrArgG1 6422
    |||
225 AATGCGCGCATCCATCGCGCGCAAGCGGTACTTCAGTCGCTC.... 270
    |||
6422 yAlaGlnAlaIleGlyAlaGlyAspValSerValAlaGlnAlaProLeu 6439
    |||
271 .....GTGATTCGCGTTGA... 285
    |||
6439 euGlyLeuGlyArgValIleAlaLeuGlnHisAlaGlnLeuArgCysAla 6455
    |||
286 .....GGCAAC...GACGAATGCA 302

```



```

;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Pyridine nucleotide transhydrogenase, subunit
; ORGANISM: A
US-08-801-344-8
```

alignment_scores:

```

Quality: 93.00 Length: 484
Ratio: 0.441 Gaps: 20
Percent Similarity: 43.595 Percent Identity: 19.835
```

alignment_block:

```
US-09-303-518D-127 x US-08-801-344-8
```

```
Align seg 1/1 to: US-08-801-344-8 from: 1 to: 510
```

```

100 GAAGATATGCGGTATGCGCCCTGATGAAATGCAAGAGCGGATGC 149
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46 Aspaaplysalaphevalglalaglialgluilevalgluglyasnse 62
150 CGTCAAAAAGGCCAAGTCTGTTGAAACAAAAAGNATCCGGCGCTGG 199
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 rValtrpInserguileileuyls..... 71
200 TGTTCACGCGCCGCTTTCAGGCAAAATCGCCCATCATCGGCC... 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
72 ..ValasnalaproleuaspaspilulealeuleuasnProglYthr 87
246 ..... 246
88 ThrleuValserPheiletrProalaglInasnProgluleuMetglInly 104
247 .....GAAAGCGCGTACTTTCAGTGCCTGATTCGCGTGAAGCA 289
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 sleuAlaglInargasnVal.....ThrValMetalaMetaserv 118
290 ACGACGATTCGAGTTCGACGCTAC.....GGCCCGCAAGCTTGCCA 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 alProargIleSerarAlaglInSerleuaspAlaleuSerSerMetala 134
334 AACTTAAGCGCGGANGAANTNNGNCAATCTGATCCAA..... 372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 AsnilealaglYtrArgalailevalglualaiahlsIuglupheglYar 151
373 .....TCGCGTTTGTGACTGCGTGCCTGACCCGCTTCAGCAAAA 415
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 gPhepHeThrleYglInlethrAlaAlaglYlYsValProAlaLysV 168
416 TCCTCGCGCTCGAT.....GCC 432
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
168 alMetValileglYalaglYalalaglYleuAlaiahlsIuglYalAla 184
433 GAGCGCTTGGCATTCGTCATGCGATGCGACCAATCCGCTNGCGC 482
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 AsnSerleuaglYalilevalargAlapheaspThrargProgluVally 201
483 AGACCTGTG.....GTTCATCAAG 505
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 sgInglInValglInSerMetglYalaglupheleuGluleuaspPheLysg 218
506 AAGCGCGANGATTCAGACGANGTTCGCTATTCAGCCGTTTG... 552
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 IuglualaglYserglYaspglYTralalalYsValmetSeraspAlaphe 234
553 .....ACGAGCGTAATAATCATGTGTGTGAAGCAGCTGGCGCAGAC... 594
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
235 IleYsalaglImetglInleuPheAlaAlaglAlalalYsgluValaspIl 251
```

```

595 .....GNGCGTCTGAAATGCTGCCACATGGA 625
251 eileValThrThrAlaleuileProglYlYsProAlaProlyleuileT 268
626 CACATGAATTCGGCGCGCCGATCCGCGCT..... 657
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 hrarglImetValaspSerMetYsalaglYserValileValaspLeu 284
658 .....TTGAGTGGCAGCAGCATTCATTCATTCAGCCG.....GT 692
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
285 AlAlaglInasnglYlYasnCysgluTrpThrValProglYgluileph 301
693 CGGTCAACAAAAACGTTTGACCAATTCATTCAGATGTAATTCGCA 742
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 eThrThrGlInasnglYalYalYsalileglYlYrThraspLeu..... 315
743 TCGGACGTTTGTTCGACAGCGCTCTGAAACCGCAGCGCGTATGCT 792
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 .....ProglYargleuProthrGlnSerSerGlnleu 326
793 TTGGTGTGTTCTCAGTCACAAACACGCGCTTTCGTCACGTTTGGG 842
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 TyrGlYThrasnleuValasn.....LeuileYsalaleuYsly 340
843 TCGCAAGATTCGCAATTCAGTCCGCGCAATTCGTCAGCAGACAC 892
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 sgluYlYsaspGlyAsnlethr.....ValaspheaspaspY 353
893 GCGTATTCGCGT.....TCGATTCGACGCGCGATTCACAGACGC 936
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 alValilearglYalThrValileargAlaglYlYlethrTrpPro 369
937 GCGCAGATTCATTCGAGCTACCAATTCAGATTCGTTATTCGAGA 986
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 AlaProProle.....GlnValSerAlaglInProgl 380
987 AGCGCCGACCAAGAG..... 1002
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
380 nAlaAlaglInlyAlaAlaProgluValYlYsThrGluYlYsCysThrC 397
1003 .....CTGTTTC 1008
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 YSerProThrparglYsTralaleuMetAlaleuAlaileleleuphe 413
1009 GGTGGGTGGCGCGCGCGAGCGGCAAAATACCTCAGCAGCGTACAGC... 1056
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 GlyTrpMetAlaSerValAlaProlysgluPheleuGlyHisPheThrVa 430
1057 .....CTCGGCC 1063
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 lPheAlaleuAlaCysValValglYlYrTrpValValTrpaspnValSerH 447
1064 ATTTCTGAAAAACAACCTTCACAGTTCAGCAGCGCTCACAGCT... 1110
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
447 lAlaleuHisThrProleuMetSerValThrAsnAlaileSerGlYlYle 463
1111 .....GGCGACCGCGCCATTCGTCGATTCGATTCAGACGCGCT 1151
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
464 lIeValValglY.....AlaleuLeuGlInleaglYlYsGlyTrpVa 478
1152 AATGCCCTAGACATTCGCTACCCCTGTTTGGCGATTTAAATCGTCG 1201
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 lSerPheleuSerPheileAlaValleuileAlaSerlIeasnlePheg 495
1202 GC 1203
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495 lY 495
```

seq_name: /cgn2_6/prodata/1/iaa/5B_COMB.pep:US-09-498-599-8

seq_documentation_block:

; Sequence 8, Application US/09498599
; Patent No. 6303352


```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/783,774
FILING DATE: 15-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7682-037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 907 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-783-774-2

```

```

alignment_scores:
  Quality: 91.00      Length: 388
  Ratio: 0.569        Gaps: 17
  Percent Similarity: 41.237      Percent Identity: 21.134

```

alignment block:
US-09-303-518d-127/rev x US-08-783-774-2 ..

Align seg 1/1 to: US-08-783-774-2 from: 1 to: 907

```

1133 ATCGGACCATGGCGCGTCCGACCGTTGAGC.....GC 1099
      |||||
396 LeuGlyThrAlaProLysThrLeuIleIleThrArgThrAlaThrAsnAl 412
      |||||
1098 TGTCGTGAACCTGAGAGTTGTTTTCAGAAATGCGGAGGTCGTAC 1049
      |||||
412 arThrThrThrHisLysValIlePheSerLysAlaProGlySerThrT 429
      |||||
1048 GCCTGATGAGATTTGTCGGCGTGGGGGCAACCCAGCGAACACSTCT 999
      |||||
429 hrThrSerProThrLeuAsnThrGlyPheAlaAspProAsnThrThr 445
      |||||
998 TTGCTGGCGCCTTTCGATACGGAATCTGATTGTGAGCTCCCAA 949
      |||||
446 ThrGlyLeuProSerSer..ThrHisVal..... 454
      |||||
948 ATATATCGTGGCGCCTTGTGTAATCGCCCGCTCAATACGAAACGGAAA 899
      |||||
455 .....ProThrAsnLeuThrAlaProAlaSerThrGlyProThrTy 468
      |||||
898 TCAGCGCGTGTGTCGCTCAACCAATCGCCGCGAGTAATTTGCGATACT 849
      |||||
468 alSerThrAlaAspValThrSerProThrProAlaGlyThrThrSerGly 484
      |||||
848 TTCGCACCCAAACGGTAGCGAAGAG.....CGTGGTTTGTTCAC 808
      |||||
485 AlaserProValThrProSerProSerProThrPaspAsnGlyThrGly 501
      |||||
807 TTGAGAACCCCAAGCAATACCGCGCGGTGTGAGCGGCGCTGTTG 758
      |||||
501 rLysAlaProAspMetThrSerSerThrSer.....ProValT 514
      |||||
757 CAACCAAAACGTCCGATGCAATTCATTGATTAATGATGTCACAAAG 708
      |||||
514 hrThrProThrProAsnAlaThr..... 521
      |||||
707 GTTTGTGTTGACCGACCGCTCATGAATGAATGATGCGTGCACCTCAA 658
      |||||

```

```

522 .....SerProThr..... 524
657 ACCGGCGGATGGGGCGCGCAATTCATGTTTCGATGTTGCAGCAT 608
      |||||
525 .ProAlaValThrThrProThrProAsnAlaThrSerProThrProAla 541
      |||||
607 TTTCAGCGGCGACGTCTGCGCCAGTGCCTTACACATGATGATTTACGC 558
      |||||
541 alThr.....ThrProThrProAsnAla..... 548
      |||||
557 TCAGTCAACGGCTCAATACACCAACATCGTCAATATCNCNCGCGC 508
      |||||
549 .....ThrSerProThrLeuGlyLysThrSerProThr 559
      |||||
507 TTCTTGTATCAACAC...ACAGGCTCTGCGCGNACGGATGTTGTC.... 465
      |||||
559 rSerAlaValThrThrProThrProAsnAlaThrSerProThrLeuGly 576
      |||||
464 .....TTCATGCAATTGACGAAGATGGCGAAGCGCTGGCA 429
      |||||
576 yThrSerProThrSerAlaValThrThrProThrProAsnAlaThrSer 592
      |||||
428 TCAGCGCGAGGATTTGCTGACGAGCGAGTACGACGCGACATCCACAA 379
      |||||
593 ProThrLeuGly..... 596
      |||||
378 ACCGATTTGATCAGATTGCGNCCNANTTCGCGCTTAAGTTTGCCA 329
      |||||
596 ..... 596
328 ACGCTTCGGCGCGTAGCGTTGCAATCGATTGCTGTCCTTCACAG 279
      |||||
597 .....LysThrSerProThrSer 602
      |||||
278 GCATTCACGACGACTGAGTACGCGCTTTCGCGCGATGATGCGCGC 229
      |||||
603 AlaValThrThrProThrProAsnAlaThrGlyProThr..ValGlyG 618
      |||||
228 GATTTTGCTGAACNGCGCGGTAAACACCAACCGCGGATNCTTTTGT 179
      |||||
618 uThrSerProGlnAlaAsnAlaThrAsnHisThr.....LeuG 631
      |||||
178 CTTCACACGACACTGGCGCTTTTTCGACGCGCATGCGCTCTCTACATT 129
      |||||
631 LysLysThrSerProThrProValAlaThr..... 640
      |||||
128 ATCNCNGGGCGCATACCGGCAATATCTCGCCCAAGCAACGCGACTG 81
      |||||
641 .....SerGlnProLysAsnAlaThrSerAl 649
      |||||
80 .GTAAATGACGGCGCGCTCATTAATGACTGTCTCGCGTCCCGCGATG 32
      |||||
649 aValThrThrGlyGlnHisAsnIleThrSerSerSerThrSerSerMet 666
      |||||
31 GCAGGTTTAGACT 18
      |||||
666 er...LeuArgPro 669
      |||||

```

seq_name: /cgn2_6/ptodata/1/lae/PCTUS_COMB.pep.PCT-US95-04611A-19

```

seq_documentation_block:
; Sequence 19, Application PC/TUS9504611A
; GENERAL INFORMATION:
; APPLICANT: Spaele, Richard and Jackman, Winthrop, T.
; TITLE OF INVENTION: Non Splicing Variants of gp350/220
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Cooley Goddard Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306

```

121 CHAACAACGTCGATGGCAATTACATCTGATAATTGATGGTCCAAACG 708

```

514 hTThProThrProAsnAlaThr..... 521
707 GTTtGTTGGACCGACGACCGCTCAATGAATGATtGCGTGCACtCA 658
522 .....SerProThr..... 524
657 ACCGGCCGGATGCGGCCCGCAATTCATGtGTTTCGATGTTGGACAT 608
525 .ProAlaValThrThrProThrProAsnAlaThrSerProThrProAlaV 541
607 TTTAGAGCGAGCTtGCGCGACCTCCCTTACACATGATtTTACGC 558
541 aThr.....ThrProThrProAsnAla..... 548
557 TCGGTCAACAGGCTCAATACCGACANtCGTGAATCtGNCGCGC 508
549 .....ThrSerProThrLeuGlyLysThrSerProTh 559
507 TTTCTTTATCAACAAC...ACAGGTCtGCGCGCAGGATtGGTG... 465
559 rSerAlaValThrThrProThrProAsnAlaThrSerProThrLeuGlyL 576
464 .....TCATCGATtGTCGAGAAGTGGCGAAGCGCTGGCA 429
576 ystThrSerProThrSerAlaValThrThrProThrProAsnAlaThrSer 592
428 TCGACGGCAGGATtTTTGCTGAACGGACGAGtACGACGCGCATCCACA 379
593 ProThrLeuGly..... 596
378 ACCGGATtGGATCAGATtGCNCtNNtNtTtCtGCGCTTAAGtTTGCCA 329
596 ..... 596
328 ACgCTTGCGGCGGTAGCGGTGAACtCGATtTCGtCGTtGCTTCAAG 279
597 .....LysThrSerProThrSer 602
278 GCATACACGACGACtGAAGtACGCGCTTTTGCCGCGATGGATGGCGGC 229
603 AlaValThrThrProThrProAsnAlaThrGlyProThr...ValGlyG1 618
228 GATTTGCCCTGAACGCGCGGTAAACACACGCGCGGATtNCtTTTGT 179
618 uThrSerProGlnAlaAsnAlaThrAsnHisThr.....LeuG 631
178 CTTCGAACACGACCTtGGCCCTTTTtTtTtTtTtTtTtTtTtTtTtTtT 129
631 LyG1yThrSerProThrProAlaValThr..... 640
128 ANtCNCGGGCGCATACCGGCATATtCTtCCGCAACACGCGACTtGC. 81
641 .....SerGlnProLysAsnAlaThrSerAl 649
80 .GTAAATGACGGCGCGCTCATTAATGACTTtGCTCCGCGTtGGCCGGATGG 32
649 aValThrThrThrGlyGlnHisAsnAlaThrSerSerSerThrSerSerMetS 666
31 GCAGGTTTAGACCT 18
666 er...LeuArgPro 669

seq_name: /cgn2_6/prodata/1/iaa/5a_comb.pep:US-08-127-499A-26
seq_documentation_block:
; Sequence 26, Application US/08127499A
; Patent No. 5510264
; GENERAL INFORMATION:
; APPLICANT: VAN ALSTYNE, Diane
; APPLICANT: SHARMA, Lawrence Rajendra
; TITLE OF INVENTION: HOMOLOGOUS WHICH BIND MENINGITIS RELATED
; TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES

```

NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30.
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/127,499A
 FILING DATE: 28-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 51916/102/INBI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 484 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 US-08-127-499A-26

alignment_scores:
 Quality: 88.00 Length: 347
 Ratio: 0.518 Gaps: 17
 Percent Similarity: 48.991 Percent Identity: 20.173

alignment_block:
 US-09-303-518D-127 x US-08-127-499A-26

Align seg 1/1 to: US-08-127-499A-26 from: 1 to: 484

```

145 GATGCCCTCAAAAAGGCCAAGTCGTGTTGAAGACAAAAGNATCCGGG 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 AspAlaIleLysLysAlaAsnAsnLeuThrThAspLysIleValProG1 66
195 CGGGTGTATTACCGCGCGCGCTTCAGGCAAAATCGCGCCATCCATCGCG 244
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 yGlu.....LysLeuGlnValAsnAsnGluValAlaIleLagLulysT 81
245 GCGAAAAGCGCGTACTT..... 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 hrgIuLysSerValSerAlaThrThrIleuAsnValArgSerGlyAlaGly 97
262 .....CAGTCGGTCGTGATTCGCGTTGAAGCGAAGCAAAATCGAGTT 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 ValAspAsnSerIleIleThrSerIleLysGlyIleThrLysValIleTrva 114
306 CGAAGCTAGCGCGCGCGAAGCGTTGGCAAACTTAAGCGGGGANGANTNN 355
|||:|||||:|||||:|||||:|||||:|||||:|||||:
114 IGLu..... 115
356 GNNCAATTCATCCATCCGGTGTGGGACTGGCTGCGTANCCGTCGG 405
|||:|||||:|||||:|||||:|||||:|||||:|||||:
116 .....ThrThrIleuSerAsnGlyTrp..... 122
406 TTCAGAAAATCCTCGCGGTGATGCGAGCGGCTTCGCAATCTTCGTCAA 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 ...HisLysIleThrTyAsnAspGlyLys.....ThrGlyPheValAs 136
|||:|||||:|||||:|||||:|||||:|||||:|||||:
456 TGGCATGAGC...ACCAATCCGCTGGCGGAGACCGCTGTGGTGTGATCA 502
|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

136 ngIuLysTyLeuThrAspLysAlaValSerThrProValAlaProThrG 153
503 AAGAAGCCGCGANGATTTAGACGANGNTGCTGATTTAGACCCCTTG 552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 InguValLysLysGlu.....ThrThrThrInguAlaAlaProAla 167
553 ACCGAGCGTAAATTCATGTCGTGAAGCGAGTCGCGAGCTGCGCTC 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 AlaGluThrLysThrGluValLysGlnThrThrGlnAlaThrThrProAl 184
603 TGAATATGCTGCCAATCGAAGCAACATGAATTGGCGCGCCGATCCG 652
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 aproLysValAla.....GluThrLysGlu.....ThrProValVala 197
653 CCGGTTGAGTGGCAGCCACATTCATTCATGTGAGCGCGGTGGCAAC 702
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 spGlnAsnAlaThrThrHis.....AlaValLysSerGly 208
703 AAACCGTTGGACCATCAATTAATCAAGATGAATTCGATCGAGCGTT 752
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 AspThrIleThrPalalaSerValLysTyGlyValSerValGlnAspI 225
753 GTTGCACAACGCGCGTCTGAACACCGCGCGTATTGCT.....T 793
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 emeSerTrpAsnAsnLeuSerSerSerSerIleTyValGlyGluLysL 242
794 TGGGTGCTTCACAGTCACAAACCAACCGCGCTTCG..... 829
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 eu.AlaIleLysGlnThrAlaAsnThrAlaThrProLysAlaGluVal 258
830 .....GTACCGTTTGGGTGCGAAGATATCGAAATTAAGTGGGGCGA 872
|||:|||||:|||||:|||||:|||||:|||||:|||||:
258 sThrGluAlaProAlaAlaGluLysGlnAlaAlaProValValLysGlu 275
873 ATGCTGACCGACAGACAAACCGCGTATTCGCGTTCGATTTGAACGG 922
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 sn.....ThrAsnThrAsn.....ThrAla 281
923 CGATTACACAAGCGCGCGCATTTATTTGGAGCTACACATCAGATT 972
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 ThrThrGluLysLysGlnThrAlaThrGlnGlnGlnThrAlaProLys 298
973 TCCG.....TTATGAGAAGAGCGCGCAAGAGCTGTTGCGCTGGCT 1016
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 aproThrGluAlaAlaLysProAlaProAlaProSerThrAsnThrAsna 315
1017 TCCGCGCGACCGCGCAAAATCTCATCAGCGGTACGACCGCTCGGCCATT 1066
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 laAsnLysThrAsnThrAsnThrAsnThrAsnThrAsn..... 326
1067 TCCTGAAAACAAACTCTTCAAGTTACAGACAGCGCGTCA 1105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 .....AsnThrAsnAsnThrAsnThrAsnThrProSer 337

```

seq_name: /cgn2_6/prodata/1/aa/5A_COMB.pep:US-08-482-847-26

seq_documentation_block:
 : Sequence 26, Application US/08482847
 : Patent No. 5556757
 : GENERAL INFORMATION:
 : APPLICANT: VAN ALSTYNE, Diane
 : TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
 : TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
 : TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THEREO, AND USES THEREOF
 : NUMBER OF SEQUENCES: 40
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W., Suite 500
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/482,847
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/127,499
 FILING DATE: 28-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 51916/104/INBT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 484 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 US-08-482-847-26

alignment_scores:
 Quality: 88.00 Length: 347
 Ratio: 0.518 Gaps: 17
 Percent Similarity: 48.991 Percent Identity: 20.173

alignment block:

US-09-303-518d-127 x US-08-482-847-26 ..

Align seq 1/1 to: US-08-482-847-26 from: 1 to: 484

145 GATGCGCTCAAAAAGGCGAAGTGTGTAAGACAAAAGATCCGGG 194
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 50 AAGPAllellyslasnasnleuThrThrasplysillevalPro1 66
 195 CGTGGTGTTCACGCGCNGTTTCAGGCAAAATCCGCCATCCACGGG 244
 -|||||:|||||:|||||:|||||:|||||:|||||:|||||
 66 ygin.....LysleuGlnValasnsglValalaalaglulysT 81
 245 GCGAAAAGCGCTACTT..... 261
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 81 hrcGlnysSerValSerAlaThrTrpLeuasnValargserglyAlagly 97
 262CAGTCGCTGTGATTCGCGTTGAAGCAGCAGCAATCGAGTT 305
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 98 ValAspAsnSerIleIleThrSerIlelysglyThrLysValThrVa 114
 306 CGAAGCTTACGGCCCGGAGCGTGGCAAACTTAAGCGCGANGAANTNN 355
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 114 lglu..... 115
 356 GNNCGCATGTGATCCAAATCGGTTTGACTGCGCTCGTANCCGTCGG 405
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 116ThrThrIuSerAsnnglyTrp..... 122
 406 TTCACCAAAATCCCTGCGGTGATGCGGCGCGTTCGCAATCTGTCAA 455
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 123 ...HisLysIleThrTyraAsnspolyls.....ThrIglPheValas 136
 456 TCGCATGAGC...ACCAATCCGCTNGCGCAGACCCCTGTGTTGATCA 502
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 136 nelysTyrlleuThrAspLysAlaValSerThrProValAlaProthrg 153
 503 AAGAGCCGCGANGATTTTCAGACGANGCTNMGCTGGATTTGAGCCGTTTG 552
 :|||:|||||:|||||:|||||:|||||:|||||:|||||

153 IncluVallyslslysglu.....ThrThrThrGlnGlnAlaAlaProAla 167
 553 ACCGAGGTAAATATCCATGTGTAGCAGCTGGCGAGCGAGTGGCGNC 602
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 168 AlaGluThrLysThrLysGlnThrThrGlnAlaThrThrProAl 184
 603 TGAATAATGCTGCCACATCAATCAATCAATTCGGCGCCGATCCGG 652
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||
 184 aProlyValAla.....GluThrLysglu.....ThrProValAla 197
 653 CCGGTTTGAGTGGCAGCAGCATTCATTCATTGAGCCGCTCGCTCAAC 702
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 197 spGlnAsnAlaThrThrHis.....AlaVallysergly 208
 703 AAACCGTTTGGACATCAATTCATCAATTCATTCATTCATTCATTCAT 752
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 209 AspThrIleTrpAlaLeuSerVallyslslyValSerValGlnAsp1 225
 753 GTTTGCAACAGCCCGCTGACACACCGAGCGGTGATTCCT.....T 793
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||
 225 emetSerTrpAsnAsnLeuSerSerSerIleTyValGlyGlnlysl 242
 794 TGGGTGCTTCACAGTCAACACACCGCTCTTTC..... 829
 ||:|||||:|||||:|||||:|||||:|||||:|||||
 242 euAlaIleLysGlnThrAlaAsnThrAlaThrProLysAlaGluVally 258
 830GTACCGTTTGGTGGCGAAGTATTCGCAATTCATTCGCGCGA 872
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 258 sThrGlnAlaProAlaAlaGluLysGlnAlaAlaProVallyslglu 275
 873 ATTGGTTGACGACGACCAACCGCGTATTCGCGTTGGTATTCGACGGC 922
 ||:|||||:|||||:|||||:|||||:|||||:|||||
 275 sn.....ThrAsnThrAsn.....ThrAla 281
 923 CGATTACACAGCGCGCAGCATTCATTGGAGCCGTACCAATCAATCAAT 972
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 282 ThrThrGluLysLysGlnThrAlaThrGlnGlnGlnAlaProLysAl 298
 973 TCCG.....TTATCGAAGAGCGCGCAGCAAGAGCTGTGCGCTGGGT 1016
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 298 aProThrGlnAlaLysProAlaProAlaProSerThrAsnThrAsn 315
 1017 TGGCGCGACCGCGACAAATCTCCATTCACGCGGTACGACCCCTGGCATT 1066
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||
 315 laAsnLysThrAsnThrAsnThrAsnThrAsnThr..... 326
 1067 TCCTGAAAACAACTCTTCAGTTCACGACGACGCGCTCA 1105
 :|||:|||||:|||||:|||||:|||||:|||||:|||||
 327AsnThrAsnAsnThrAsnThrAsnThrProSer 337

seq_name: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-907-166-10

seq_documentation block:

; Sequence 10, Application US/08907166
 ; Patent No. 5948666
 ; GENERAL INFORMATION:
 ; APPLICANT: Callen, Walter
 ; APPLICANT: Mather, Eric
 ; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
 ; FILE REFERENCE: 09010/027001
 ; CURRENT APPLICATION NUMBER: US/08/907,166
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 10
 ; LENGTH: 762
 ; TYPE: prt
 ; ORGANISM: Desulfurococcus sp.
 ; FEATURE:
 ; NAME/KEY: variation
 ; LOCATION: (601)..(601)
 ; OTHER INFORMATION: Xaa at position 601 is alanine or proline
 US-08-907-166-10

alignment_scores: Quality: 87.50 Length: 400
 Ratio: 0.494 Gaps: 15
 Percent Similarity: 44.250 Percent Identity: 19.750

alignment_block:
 US-09-303-518D-127 x US-08-907-166-10 ..

Align seg 1/1 to: US-08-907-166-10 from: 1 to: 762

```

169 CTGTTTGACGACAAAAGNATCCGGCGGTGTTTACCCGCCGCTTTG 218
      |||
39  ILeYrAspAsnArgLeuIleProGlyMetArgYrValThrAspPhe 55
      |||
219 AGGCAAAATCCGCCCATCCATCCGGCGGAAAGCGGCTTCTACTGCG 268
      |||
55  rAsnGlyAlaGlnLysLeuValMetValLysProGluIleProGlnSerL 72
      |||
269 TCGTGATTGCCGTTGAAGCAACGCAATCGATTGCAACGCTACCGC 318
      |||
72  euValGluLysValArg.....GluLeu...PheArgAsnGluPro 84
      |||
319 CCCGAAGCGCTGGCAACTTAAGCGCGGAGNAAATNNGNCCATCTGAT 368
      |||
85  ProGluThrVal.....LysLeuAl 91
      |||
369 CCAATCCGGTTTGTGACTGGCTGCTGACGCTCCGCTTCAGCAAAATCC 418
      |||
91  aGluGluLeuLeuLeuPheGluGluSerProProLysValLysArgV 108
      |||
419 CTGCGCGTGAATCCGACCGCTTCCGCAATCTGTCATACGATGACAC 468
      |||
108 aLaValAlaSpIleGluValPheThrProPheLysGlyArgIleProSer 124
      |||
469 AATCCGCTNGCGGACAGACCCCTGGTGTGATCAAGAAGCCGNGANGA 518
      |||
125 ProLysLeuAlaGluTyrProValIleSerIleAlaLeuAlaGlySerAs 141
      |||
519 TTTCAAGACAGANTTNGCTGATTAAGCCGTTTGACCGGCTTAAATCC 568
      |||
141 pGlyLeuLysLysIleLeuLeuAlaArg.....GluTyrLys...H 155
      |||
569 AATGTGTGAAGGAGCTGGCGGAGAGCTCCG..... 600
      |||
155 IAspAspPheAspTyrMetMetGluAspTyrProValGluAlaGluValGlu 171
      |||
601 .....TCTGAATATGCTGCCAATCAATCAACATGAAATTCGGCGG 641
      |||
172 ValPheAspSerGluLysAspMetLeuLeuAlaPheArgIleMetGlu 188
      |||
642 CCGGCATCCGGC..... 654
      |||
188 ySerTyrProValValLeuThrTyrAsnGlyAspAsnPheAspLeuGluT 205
      |||
655 .....GTTTGAGTGCACGACATTCAT 678
      |||
205 yLeuTyrValAlaGlnAlaPheLysLeuGlyIleLeuArgSerHisIle... 220
      |||
679 TTCATTTAGCGCGGTGCGTGCACAAACCGTTTGACCATCAATTAATCA 728
      |||
221 .....ProLeuLysIleGlyGluAspMetIleArgIleAspPhe 234
      |||
729 AGATGTAATTCGATCGAGCTTGTGTCACA..... 762
      |||
234 rIleHisLeuAspLeuTyrLysPhePheSerAsnArgAlaValLysAsnT 251
      |||
763 .....GGCGCTGACACCGGCGCTGATTTGCTTTGGTGGT 801
      |||
251 yAlaLapheGlyLysTyrGlnGluLysLeuAspAlaValSerGly 267
      |||
802 TCT.....CAATCAACAAACACGCGCTTGTGCTACCGTTTGGGTGC 845
  
```

```

268 AlaleuLeuGlyValSerLysIleGlyPheGluGluThrIleGlyLyl 284
      |||
846 GAAGTATCCGAATTTACTCGGGCGAATGTTGACGCGACACACCGCG 895
      |||
284 eSerAlaSerLeuLeuAlaAlaTyrAsnTyrArgAspAlaGlu..... 298
      |||
896 TGATTTCCGGTTCGATTGACGCGCGCATTTACACAGGCGCGCAGAT 945
      |||
299 .....IleThrLeuAsn..... 302
      |||
946 TATTTGGACGCTACCAATCAGATT..... 972
      |||
303 ...LeuThrMetPheSerAsnGluLeuValTyrLysLeuIleIleLeu 318
      |||
973 .....TCCGTTATCGAAGAAGCGCGCAGCAGCAGACGCTG 1006
      |||
318 uAlaArgValSerLysThrSerIleGluAspLeuLysArgArgGlnIle 335
      |||
1007 TCGGCTGGGTTGCCCGCAGCGGCAAAATCTCCATCAGCGCTACGACC 1056
      |||
335 eTyrTyrPhe.....GlnAsnLeuPhePheTyrGluArgArgLys 348
      |||
1057 CTGCGCATTTCTCGAAAACAACTTTCAGTTCACGACCGCCGCA 1106
      |||
349 LeuGlyTyrLeuIleProAsnLysGluAspIleLeuArgHisValArgI 365
      |||
1107 CGGTGGCGCGCGCGCATGTCG..... 1128
      |||
365 yThrGlyThrLysAlaIleIleGlyGlyLysTyrAlaGlyAlaLeuV 382
      |||
1129 .....CCGATTTGTACTTACGAGCGCGTAATGCGCGCTAGCATC 1167
      |||
382 aValGluProProLysGlyAlaPhePheAsnValAlaValLeuAspIle 388
      |||
seq_name: /cgn2_6/ptdata/1/1aa/6b_COMB.pep:US-09-351-200-2
  
```

```

seq_documentation_block:
; Sequence 2, Application US/09351200
; Patent No. 6320033
; GENERAL INFORMATION:
; APPLICANT: BOURBONNAIS, Yves
; APPLICANT: LAMARRE, Claude
; TITLE OF INVENTION: CANDIDA ALBICANS GENE (CSA1) ENCODING A
; FILE REFERENCE: 6013-71"US" CC/
; CURRENT APPLICATION NUMBER: US/09/351,200
; EARLIER FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: CA2,237,134
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (0)...(0)
US-09-351-200-2
  
```

alignment_scores: Quality: 87.50 Length: 334
 Ratio: 0.534 Gaps: 10
 Percent Similarity: 49.102 Percent Identity: 20.958

alignment_block:
 US-09-303-518D-127/rev x US-09-351-200-2 ..
 Align seg 1/1 to: US-09-351-200-2 from: 1 to: 1203

221 yThrileProcysasplysPheProSer 230

seq_name: /cgn2_6/prodata/1/iaa/5B_COMB.pep:US-08-396-001-14

seq_documentation_block:

Sequence 14, Application US/08396001

Patent No. 5919618

GENERAL INFORMATION:

APPLICANT: Guarente, Leonard P.

APPLICANT: Austriaco Jr., Nicanor

APPLICANT: Claus, James

APPLICANT: Cole, Francesca

APPLICANT: Kennedy, Brian

TITLE OF INVENTION: Genes Determining Cellular Senescence in

NUMBER OF INVENTION: Yeast

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/396,001

FILING DATE: 28-FEB-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: MIT-6408a2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 475 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-396-001-14

alignment_scores:

Quality: 87.00 Length: 226

Ratio: 0.784 Gaps: 6

Percent Similarity: 49.115 Percent Identity: 22.566

alignment_block:

US-09-303-518d-127/rev x US-08-396-001-14

Align seg 1/1 to: US-08-396-001-14 from: 1 to: 475

764 CCTGTGCAACAAACGTCGATGCAATTCATCTTGATATGATGAT 715

715 ProAlaIaSerValLeuAlaIaProIhSer.....AlaIleI 71

714 CCAACGGTTTGTTCGACCGACCGGCTCAATGAATGATGCGGTC 665

71 ethrIhSerValLeuAlaIaProIhSer..... 81

664 CACTCAACCGCGCGATGCGGCGCGCAATTCATGTTGATGTTG 615

82AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 96

614 GCAGCATTTTCAGACGACGCTGTGCGCGCACTGCTTACACATGAT 565

97 LeuAlaIaSerValLeuAlaIaProIhSer..... 108

564 TTTACGCTCGTCAACAGCGCTCATACACAGCANACGCTTGAATGAT 515

109SerAlaIhIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 116

514 CGNCGCTTCTTGTATCATACACAGCGCTGCGCANACGATGATG 465

116 IndIyAlaSerSerSerSerSerSerSerSerSerSerSerSerSer 132

464 TCCATGCAATTCAGACAGATGCGCAACGCTTGGATGACGCGCAG 415

133 SerSerSerValSerSerSerSerSerSerSerSerSerSerSerSer 149

414 TTTGCTGAACGACGAGGTACGACGCGCAGTCCACAAACCGATGATCA 365

149 Val..... 150

364 GATTGCGNCGNANTCNCGCGCTTAAGTTTGCACAGCTTGGCGCG 315

151SerThrSerSerAlaIhIhIhIhIhIhIhIhIhIhIhIhIhIhIhIhIh 163

314 TAGCGTGCACACTGATTCGTCGTCCTTCAACGCGCATACAGACCGA 265

164 ThrIySerSerSerSerSerSerSerSerSerSerSerSerSerSer 180

264 CTGAAGTACGCGCTTTCGCGCGATGATGCGCGATTTGCTGAA 215

180 ThrSerSerSerSerSer.....SerSers 188

214 CNGGCGGCTAAACACACGCGCGGATTCCTTTGCTTCAACAGCACT 165

188 erSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 204

164 TGCGCTTTTGTAGC.....GCATCGCTTCTTGAATTCACACNAGG 121

205 TyrGIa 221

120 GCGCATACCGGCAATTCCTTCCCAAGC 93

221 yThrileProcysasplysPheProSer 230

seq_name: /cgn2_6/prodata/1/iaa/5B_COMB.pep:US-09-323-433A-14

seq_documentation_block:

Sequence 14, Application US/09323433A

Patent No. 6218512

GENERAL INFORMATION:

APPLICANT: Guarente, Leonard P.

APPLICANT: Austriaco Jr., Nicanor

APPLICANT: Claus, James J.

APPLICANT: Cole, Francesca

APPLICANT: Kennedy, Brian

TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESCENCE IN

FILE REFERENCE: 0050.1491-003

CURRENT APPLICATION NUMBER: US/09/323,433A

PRIOR FILING DATE: 1999-06-01

PRIOR APPLICATION NUMBER: US 08/396,001

PRIOR FILING DATE: 1995-02-28

PRIOR APPLICATION NUMBER: PCT/US94/09351

PRIOR FILING DATE: 1994-08-15

PRIOR APPLICATION NUMBER: US 08/107,408

PRIOR FILING DATE: 1993-08-16

NUMBER OF SEQ ID NOS: 48

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 14

LENGTH: 475

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-09-323-433A-14

alignment_scores: 87.00 Length: 226
 Ratio: 0.784 Gaps: 6
 Percent Similarity: 49.115 Percent Identity: 22.566

alignment_block:
 US-09-303-518D-127/rev x US-09-323-433A-14 ..

Align seg 1/1 to: US-09-323-433A-14 from: 1 to: 475

```

764 CCGTGTGCAACCAACGTCGATTCATGATTAATGATGCT 715
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 ProlaAlaSerGluValAlaThAlaAlaThSer.....AlaIleI 71
|||:|||||:|||||:|||||:|||||:|||||:|||||:
714 CCAACGGTTTGTGGACCGACCGGCTCAATGAATGATGCGTGC 665
|||:|||||:|||||:|||||:|||||:|||||:|||||:
71 ethrThrSerValLeuAlaProThrSer..... 81
|||||:|||||:|||||:|||||:|||||:|||||:
664 CACTCAACCGCGGATGCGCGCGCGGATTCATGCTGATGCTG 615
|||||:|||||:|||||:|||||:|||||:|||||:
82 .....AlaAlaAlaGlyIleAlaAlaSerIleAlaValSerSerAlaAla 96
|||||:|||||:|||||:|||||:|||||:|||||:
614 GCAGGATTTTGCAGCGACGCTGCGCGGCTTACACACATGAT 565
|||||:|||||:|||||:|||||:|||||:|||||:
97 LeuAlaIlyAsnGluLysIleSerAspAlaAla..... 108
|||||:|||||:|||||:|||||:|||||:|||||:
564 TTACGCTCGGTCAACGCGCTCAATACAGCAACGCTGCTGAATC 515
|||||:|||||:|||||:|||||:|||||:|||||:
109 .....SerAlaThrAlaSerThrSerg 116
|||||:|||||:|||||:|||||:|||||:|||||:
514 CCGCGCTTCTTGTATCACAACCAACAGGCTGCGCGGAGGATG 465
|||||:|||||:|||||:|||||:|||||:|||||:
116 InGlyAlaSerSerSerSerSerSerSerAlaThSerThrLeuGlu 132
|||||:|||||:|||||:|||||:|||||:|||||:
464 TCCATGCGATTTGACGAAGATGGCAACGCTGCGGATCGACGAG 415
|||||:|||||:|||||:|||||:|||||:|||||:
133 SerSerSerValSerSerSerSerSerGluGluAlaAlaProThr 149
|||||:|||||:|||||:|||||:|||||:|||||:
414 TTTGCTGAACGAGCGGNTACGACGCGCATGCCAACCAGATTG 365
|||||:|||||:|||||:|||||:|||||:|||||:
149 lVal..... 150
|||||:|||||:|||||:|||||:|||||:|||||:
364 GATTGNNNNNTTNTGCGCGCTTAAGTTGCCACGCTTGGGCG 315
|||||:|||||:|||||:|||||:|||||:|||||:
151 .....SerThrSerSerAlaThInSerSerAlaSerSerAla 163
|||||:|||||:|||||:|||||:|||||:|||||:
314 TAGCGTTGCAACTGATTCGTCGTCCTTCACGCGATCAGACG 265
|||||:|||||:|||||:|||||:|||||:|||||:
164 ThrIlySerSerThrSerSerThrSerThrSerThrSerThr 180
|||||:|||||:|||||:|||||:|||||:|||||:
264 CTGAGTAGCGGCTTTTCCCGCGGATGATGGCGGCGATTGGCT 215
|||||:|||||:|||||:|||||:|||||:|||||:
180 rThrSerSerThrSerSer.....Sers 188
|||||:|||||:|||||:|||||:|||||:|||||:
214 CNGGCGGGTAACACACCGCGGATNCTTTTGTCTCAACAGCA 165
|||||:|||||:|||||:|||||:|||||:|||||:
188 erSerSerSerSerSerSerSerSerSerSerSerSerSer 204
|||||:|||||:|||||:|||||:|||||:|||||:
164 TGCGCTTTTGTAGC.....GCATGCGCTTCTGACTTTGATC 121
|||||:|||||:|||||:|||||:|||||:|||||:
205 TyrcIlyAspLeuAlaAspSerGlyProSerGluLysPheGln 221
|||||:|||||:|||||:|||||:|||||:|||||:
120 GCGGATCGCGGATTTCTTCCGCGACG 93
|||||:|||||:|||||:|||||:|||||:|||||:
221 yThrIleProCysAspLysPheProSer 230
|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /cgn2_6/ptodata/1/iaa/5A-COMB.pep:US-08-967-513-5

seq_documentation_block:
 ; Sequence 5, Application US/08967513
 ; Patent No. 5783436
 ; GENERAL INFORMATION:
 ; APPLICANT: Robert P. Hausinger

```

? TITLE OF INVENTION: Mutant Urease and Method
? TITLE OF INVENTION: of Use For Determination
?
? NUMBER OF INVENTIONS: of Urea
?
? NUMBER OF SEQUENCES: 5
?
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Ian C. McLeod
? STREET: 2190 Commons Parkway
? CITY: Okemos
? STATE: Michigan
? COUNTRY: USA
?
? ZIP: 48864
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
? MEDIUM TYPE: storage
?
? OPERATING SYSTEM: IBM Compatible
? SOFTWARE: Wordperfect 5.1
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/967,513
? FILING DATE: 11-NOV-1997
? CLASSIFICATION: 435
?
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/687,645
? FILING DATE: July 26, 1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Ian C. McLeod
? REGISTRATION NUMBER: 20,931
? REFERENCE/DOCKET NUMBER: MSU 4.1-309
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (517) 347-4100
? TELEFAX: (517) 347-4103
? TELEX: No. 5783436e
?
? INFORMATION FOR SEQ ID NO: 5:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 567
? TYPE: amino acids
? STRANDEDNESS: Single
? TOPOLOGY: Linear
?
? MOLECULE TYPE:
? DESCRIPTION: protein
?
? HYPOTHETICAL: NO
?
? ANTI-SENSE: NO
?
? ORIGINAL SOURCE:
? ORGANISM: N/A
? STRAIN: N/A
? INDIVIDUAL ISOLATE: N/A
? CELL TYPE: N/A
?
? FEATURE:
? NAME/KEY: subunit Urec
? LOCATION:
? IDENTIFICATION METHOD: Sequencing
? OTHER INFORMATION: Encoded subunit of mutant
? OTHER INFORMATION: urease
?
? US-08-967-513-5

```

alignment_scores: 86.50 Length: 247
 Ratio: 0.832 Gaps: 12
 Percent Similarity: 42.105 Percent Identity: 23.077

alignment_block:
 US-09-303-518D-127 x US-08-967-513-5 ..

Align seg 1/1 to: US-08-967-513-5 from: 1 to: 567

```

541 TTGAGCGGTTTGCAGCGGTAATTCATGTTGTAAGCAGCTGGCG 590
|||||:|||||:|||||:|||||:|||||:|||||:
260 LeuAlaAlaIlecllyGlyArgThrIleHisThrPheHisThrGlu 276
|||||:|||||:|||||:|||||:|||||:|||||:
591 AGAGCTGCGCTGAAATGCTGCCAACAATCGAACAACATGATTCG 640
|||||:|||||:|||||:|||||:|||||:|||||:
276 a.....GlyG 278

```

```

641 GCCCGCATCCGCGCGTTTGAGTGGCAGCACATTTCAT...TTCATTGAG 687
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
278 LGLYHISAlaProaspIleIleThrIlaCysAlaHisProAsnIleLeu 294
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
688 CCGGTCGGTGCAGCAACAAACCGTT...TGCACATCAATTATCAAGATGT 734
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
295 ProSerSerThrAsnProThrLeuProTyrThrLeuAsnThrIleAspGI 311
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
735 A.....ATTGCCA 742
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
311 whIsleuaspMetleuMetValCysHisHisleuaspProaspIleIleag 328
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
743 TCGGACGCTTTTGGCAACAGCGCGTGTGAACACCGCGCGTATTGCT 792
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
328 LuAspValAlaIaPheAlaGluSerArgIleArgArgIuThrIleAlaIa 344
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
793 .....TTGGGCGGTTCACAGTGCACAAACACCGC... 822
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
345 GluAspValIleuHisaspIleuGlyAlaIaPheSerleuThrSeraspse 361
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
823 .....CTCTTGGCGTACCGTTTGGGTG 844
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
361 rGlAlaMetGlyArgValGlyIuValIleleuArgThrTyrPglValA 378
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
845 CGAAAGATATGCCAATTACTGCGCGCAATTGGTT.....GAC 882
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
378 IaHisArgMetLysValGlnArgGlyAlaLeuAlaGluIuThrGlyAsp 394
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
883 GCAGACAAAC.....CGCGATTTCCGCGTTCGGTATGTAACCG 920
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
395 AsnaspAsnPheArgValLysArgTyrIleAlaLysTyrThrIleAsnTr 411
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
921 CGCGATTACAAAGC...GCGCAGATTAATTTGGACGCTACCAACATC 967
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
411 oAlaLeuThrHisGlyIleAlaHis.....G 420
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
968 AGATTTCGTTATCGAAGAGCGCGCAAGAGCTGTTCGGCTGGGT 1017
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
420 LuValIleGlySerIleGlyValGlyLysLeuAlaAspLeuValIleTrp... 435
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1018 GCGCGCGACGCGGACAATACTGCATCAGCGTACAGCCCTGCGCATTT 1067
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
436 .....SerProAlaPhePheGlyVal..... 442
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1068 CTGGAAGAAACAAACTCTTCAGATTTCAGCAGACCGTCAACGGTGGCAGC 1117
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
443 .....LysProAlaThrValIleLysGlyGlyMetI 453
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1118 GCGCGCATGTGCGGATTGGTACTTACGAGCGCGTAAATGCGG 1158
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
453 leAlaIleAlaProMetGlyAspIleAsnAlaSerIlePro 466
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-687-645B-5

seq_documentation_block:

; Sequence 5, Application US/08687645B

; Patent No. 5846752

; GENERAL INFORMATION:

; APPLICANT: Robert P. Hausinger

; TITLE OF INVENTION: Mutant Urease and Method

; TITLE OF INVENTION: of Use For Determination

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESS: Ian C. McLeod

; STREET: 2190 Commons Parkway

; CITY: Okemos

; STATE: Michigan

; COUNTRY: USA

; ZIP: 48864

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb

```

MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,645B
FILING DATE: July 26, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-309
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 567
TYPE: amino acids
STRANDEDNESS: Single
MOLECULE TYPE:
DESCRIPTION: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: N/A
STRAIN: N/A
INDIVIDUAL ISOLATE: N/A
CELL TYPE: N/A
FEATURE:
NAME/KEY: subunit Urec
LOCATION:
IDENTIFICATION METHOD: Sequencing
OTHER INFORMATION: Encoded subunit of mutant
OTHER INFORMATION: urease
US-08-687-645B-5

```

alignment_scores:

Quality: 86.50 Length: 247
Ratio: 0.832 Gaps: 12
Percent Similarity: 42.105 Percent Identity: 23.077

alignment_block:

US-09-303-518D-127 x US-08-687-645B-5 ..

Align seg 1/1 to: US-08-687-645B-5 from: 1 to: 567

```

541 TTGAGCCGTTTGACCGAGCGTAATAATCCATGTGTGAAGCAGCTGGCGC 590
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
260 LeuAlaAlaIleGlyArgThrIleHisThrPheHisThrGluGlyAl 276
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
591 AGACGTGCGCTGTGAATAATGCTGCCAACATCGAACAACATGATTTGGCGC 640
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
276 A.....ATTGCCA 742
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
641 GCCCGCATCCGCGCGTTTGAGTGGCAGCACATTTCAT...TTCATTGAG 687
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
278 LGLYHISAlaProaspIleIleThrIlaCysAlaHisProAsnIleLeu 294
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
688 CCGGTCGGTGCAGCAACAAACCGTT...TGCACATCAATTATCAAGATGT 734
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
295 ProSerSerThrAsnProThrLeuProTyrThrLeuAsnThrIleAspGI 311
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
735 A.....ATTGCCA 742
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
311 whIsleuaspMetleuMetValCysHisHisleuaspProaspIleIleag 328
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

743 TCAGACGTTGTTTCAGACAGCCGCTGTGACACCGACCGCTGATTGCT 792
      ||||| :|||:||||| :||| :|||
328 LAApValAlaIphaIaGluSerArgIleArgArgIleThrIleAlaIa 344
      ||||| :|||:||||| :||| :|||
793 .....TTGGGGGTTTCAGTCAAGTCAACAAACACGCG... 822
345 GluAspValLeuHisAspLeuGlyAlaIphaSerLeuThrSerAspSe 361
      ||||| :|||:||||| :||| :|||
823 .....CTCTGCGTACCGCTTTTGGGCTG 844
      ||||| :|||:||||| :||| :|||
361 rGlnAlaMetGlyArgValGlyGluValIleLeuArgThrTrpGlnVala 378
      ||||| :|||:||||| :||| :|||
845 CGAAAGTATCGCAATTACTGCGGGCGAATTGTT.....GAC 862
      ||||| :|||:||||| :||| :|||
378 LAlaIArgMetLeuValGlnArgGlyAlaLeuAlaGluGlnThrGlyAsp 394
      ||||| :|||:||||| :||| :|||
883 GCAGCAAC.....CGCGTATTTCCGCGTCCGATTGTAACGG 920
      ||||| :|||:||||| :||| :|||
395 AsnAspAsnIphaArgValIleArgThrIleAlaLysThrIleAsnPr 411
      ||||| :|||:||||| :||| :|||
921 CGCGATTACACAGCG...GGCAGCATTTATTGGACGCTACACATC 967
      ||||| :|||:||||| :||| :|||
411 oAlaLeuThrHisGlyIleAlaHis.....G 420
      ||||| :|||:||||| :||| :|||
968 AGATTTCGTTATCGAGAGAGCGCGACAAAGAGCTGTCGCTGGCTT 1017
      ||||| :|||:||||| :||| :|||
420 LValGlySerIleGlyValGlyLysLeuAlaAspLeuValTrp... 435
      ||||| :|||:||||| :||| :|||
1018 GCGCCGACCGCGACAAATFACTCCATCAGCGGTACAGCCCGCCCATTT 1067
      ||||| :|||:||||| :||| :|||
436 .....SerProAlaIphaIheGlyVal..... 442
      ||||| :|||:||||| :||| :|||
1068 CCTGAAGAAACAACTCTCAAGTTCACGACGCGTCAACGCTGCGACG 1117
      ||||| :|||:||||| :||| :|||
443 .....LysProAlaThrValIleLysGlyMetI 453
      ||||| :|||:||||| :||| :|||
1118 GCGCCATGTCGCGATTGGTACTTACGAGCGCGTAATGCGG 1158
      ||||| :|||:||||| :||| :|||
453 LAlaIleAlaIphaProMetGlyAspIleAsnAlaSerIlePro 466
      ||||| :|||:||||| :||| :|||

```

seq.name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.US-08-456-670B-40

seq.documentation_block:

```

: Sequence 40, Application US/08456670B
: Patent No. 5932415
:
: GENERAL INFORMATION:
: APPLICANT: SCHUBERT, PETER
: APPLICANT: NEUMANN, STEGFRIED
: APPLICANT: PAMELZIK, MARTINA
: APPLICANT: LINKELEIER, WINFRIED
: APPLICANT: BURGER, CHRISTA
: APPLICANT: HOFMANN, GOTTFRIED
: APPLICANT: ROBERT, ANDREAS
: APPLICANT: GOEBEL, WERNER
: APPLICANT: KOHLER, STEFAN
: TITLE OF INVENTION: PROCESSES AND AGENTS FOR DETECTING
: TITLE OF INVENTION: LISTERIAS
: NUMBER OF SEQUENCES: 43
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MILLEN, WHITE, ZELANO & BRANTIGAN, P.C.
: STREET: 2200 CLARENDON BLVD., SUITE 1400
: CITY: ARLINGTON
: STATE: VIRGINIA
: COUNTRY: US
: ZIP: 22201
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/456.670B
: FILING DATE: 01-JUN-1995

```

```

: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/412,227
: FILING DATE: 27-MAR-1995
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/075,248
: FILING DATE: 11-JUN-1993
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE 4239567.4
: FILING DATE: 25-NOV-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE 421911.4
: FILING DATE: 11-JUN-1992
:
: ATTORNEY/AGENT INFORMATION:
: NAME: HAMLET-KING, DIANA
: REGISTRATION NUMBER: 33,302
: REFERENCE/DOCKET NUMBER: MERCK 1694D1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-243-6333
: TELEFAX: 703-243-6410
:
: INFORMATION FOR SEQ ID NO: 40:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 478 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: ORGANISM: Listeria monocytogenes
: STRAIN: EGD
:
: US-08-456-670B-40

```

alignment_scores:

Quality	Ratio	Length	Gaps
86.00	0.494	348	17
Percent Similarity: 50.000		Percent Identity: 20.402	

alignment_block:

US-09-303-518d-127 x US-08-456-670B-40 ..

Align seg 1/1 to: US-08-456-670B-40 from: 1 to: 478

```

145 GATGCCGCTAAAGGCCAAGTCTGTTGAACACAAAGNATCCGGG 194
      ||||| :|||:||||| :||| :|||
50 AspAlaIleLysIleAlaAsnAsnLeuThrThrAspLysIleValProG1 66
      ||||| :|||:||||| :||| :|||
195 CGTGTGTTTACCGCGCGCGTTCAGCAAAATCGCGCATCATCGCG 244
      ||||| :|||:||||| :||| :|||
66 yGln.....LysLeuGlnValAsnAsnGlnValAlaAlaIaGluLysT 81
      ||||| :|||:||||| :||| :|||
245 GCGAAAGCGCGTACTT..... 261
      ||||| :|||:||||| :||| :|||
81 hrcIuLysSerValSerAlaThrTrpLeuAsnValArgThrGlyValaGly 97
      ||||| :|||:||||| :||| :|||
262 .....CAGTGGTCTGATTGCCGTTGAAGCAACGCAAAATGAGATT 305
      ||||| :|||:||||| :||| :|||
98 ValAspAsnSerIleIleThrSerIleLysGlyThrLysValThrVa 114
      ||||| :|||:||||| :||| :|||
306 CGAAGCTACGCGCGCGGAGCGTGGCAAACTTAAGCGCGGANGAANTNN 355
      ||||| :|||:||||| :||| :|||
114 Gln..... 115
      ||||| :|||:||||| :||| :|||
356 GNNGCAATCTGATCAATCGCGTTGAGACGCGCTGCTGANTCCGTCG 405
      ||||| :|||:||||| :||| :|||
116 .....ThrThrGluSerAsnGlyTrp..... 122
      ||||| :|||:||||| :||| :|||
406 TTCAGCAAAATCCCTGCGTCGATGCGCGGCGGCTTGCATCTTGCTAA 455
      ||||| :|||:||||| :||| :|||
123 ...HisLysIleThrIleAsnAspGlyLys.....ThrGlyPheValas 136
      ||||| :|||:||||| :||| :|||
456 TGCATGATGAC...ACCAATCCGCTNGCGGACGACCGCTGTGTGATGAC 502
      ||||| :|||:||||| :||| :|||

```

```

136 nGlyLysTyrLeuThrAspLysAlaValSerThrProValAlaProThrG 153
503 AAGAGCCGCGCAGCATTTTCAGACAGANGTNGCTGTATTGACCGCTTG 552
133 lngluValLysLysglu.....ThrThrThrGlnGlnAlaAlaProVal 167
553 ACCGAGCCTTAAATTCATGTGTGTAAAGCAGCTGGCGCAGACGTGCGTC 602
168 AlaGluThrLysThrGluValLysGlnThrGlnAlaThrThrProAl 184
603 TGAATAATCTGCGCAACATCGAACATGATTTGGCGCCCGCATCCG 652
184 aProLysValAla.....GluThrLysglu.....ThrProValIleA 197
653 CCGGTTGAGTGGCAGCAGCATTCATTTTCATTTGAGCCGGTGGTCAAC 702
197 spGlnAsnAlaThrThrHis.....AlaValLysSerGly, 208
703 AAAACGCTTTGGACCATCATATTATCAAGATTAATGCCATCGACGCTTT 752
209 AspThrIleThrPalaLeuSerValLysTyrGlyValSerValGlnAspI 225
753 GTTTCGAACAGCGCGCTCGAACACCGCGCGCTGATTGCT.....T:793
225 emetSerThrPasnAsnLeuSerSerSerSerIleTyrValGlyGlnLysL 242
794 TGGGTGGTTCACAGTCACAAACACGACGCTTGC.....829
242 eu.AlaIleLysGlnThrAlaAsnThrAlaThrProLysAlaGluVally 258
830 .....GTACCGTTTGGGTGGCAAAAGTATCGCAATTACTGCGGCGA 872
258 sThrGluAlaProAlaAlaGluLysGlnAlaAlaProValVallygluA 275
873 ATTGGTTGACGACAGACACCGCGGATTTCGGTGGTTCGATTGAACGGCG 922
275 sn.....ThrAsnThrAsn.....ThrAla 281
923 CGATTACACAGGCGCGCAGATTAATTGGAGCTACACGACATCAGATT 972
282 ThrThrGluLysLysGluThrAlaThrGlnGlnGlnThrAlaProLysAl 298
973 TCCG.....TTATCGAAGAGCGCGCAGAAAGACCTGTGCGCTGGGT 1016
298 aProThrGluAlaAlaLysProAlaProAlaProSerThrAsnThrAsnA 315
1017 TGGCGCGCAGCGCAGCAATTAATCTCATCAGCGGTACGACCTCGGCGCAT 1066
315 laAsnLysThrAsnThrAsnThrAsnThrAsn...AsnThrAsnThrPro 330
1067 TCCTGAAAACAAACTCTTCAGTTCACGACGACCGCTCAACG 1108
331 SerLysAsnThrAsnThrAsnSerAsnThrAsnThrAsnThr 344
seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-487-886-2

```

seq documentation block:

```

: Sequence 2, Application US/08487886
: Patent No. 5744448
: GENERAL INFORMATION:
: APPLICANT: Kelton, Christie Ann
: APPLICANT: Schweichhardt, Rene Lynn
: APPLICANT: Cheng, Shirley Yui Yen
: APPLICANT: Nugent, No. 3744448een Patrice
: TITLE OF INVENTION: Human Follicle Stimulating
: NUMBER OF INVENTION: Hormone Receptor
: CORRESPONDENCE ADDRESS: 2
: ADDRESSEE: Stephan P. Williams,
: ADDRESSEE: Ares-Serono, Inc.
: STREET: Exchange Place, 37th floor
: CITY: Boston
: STATE: MA

```

```

: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
: COMPUTER: IBM PS/2, model 55 SX
: OPERATING SYSTEM: MS-DOS version 4.0
: SOFTWARE: VAX/VMS Massill via Kermit to IBM MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,886
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/670,085
: FILING DATE: 15-MAR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Williams, Stephan P.
: REGISTRATION NUMBER: 28546
: REFERENCE/DOCKET NUMBER: US/252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 723-1300
: TELEFAX: (617) 723-8923
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 695
: TYPE: Amino acid
: TOPOLOGY: Linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: signal sequence
: LOCATION: -17 to -1
: IDENTIFICATION METHOD: hydrophobic
: FEATURE:
: NAME/KEY: putative amino-terminal extracellular domain
: LOCATION: 1 to 349
: IDENTIFICATION METHOD: similarity with other
: IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular
: IDENTIFICATION METHOD: domains, hydrophilic
: FEATURE:
: NAME/KEY: transmembrane domain
: LOCATION: 350 to 613
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains
: FEATURE:
: NAME/KEY: putative transmembrane region I
: LOCATION: 350 to 370
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region II
: LOCATION: 382 to 404
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region III
: LOCATION: 427 to 448
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region IV
: LOCATION: 469 to 491
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region V
: LOCATION: 512 to 533
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
: FEATURE:

```



```

742 ATCGAGCTTTGTTGCAACAGCCCTCTGAC..... 774
   ::::: ||| ::::: |||
241 LeuylslyslleuargalaargserthrtyrASNleuylsleuproth 257
775 .ACGAGCGCGGTATGCTGGTGGTTCAGATCAACAAACAGCC 823
   |||::: |||::: |||::: |||::: |||::: |||:::
257 rleuglulysleuvalalaMetGluAlaSerleuthrtyrPro.Ser 273
824 TCTTCGTCACGCTTTGGGTGCGAAGATATGCAAAATTAATGCGGCGAA 873
   ||| ||| ||| ||| ||| ||| ||| ||| |||
274 Hiscyscysalaphne.....Alaas 280
874 TTGCTTGACGACGACACCGCGGTATTCGGTTCGATGACGCGGC 923
   ||| ||| ||| ||| ||| ||| ||| ||| |||
280 ntrpararglnlileserGluLeuHisProIle..Cys..AsnLysSe 295
924 GATTACACAGAGCGCGCAGATTTATTTGGACGTACCAATCAG...A 970
   ||| ::::: |||::: |||::: |||::: |||::: |||:::
295 rtleuarglngluvalasptyrmetthrGlnThrarglylnarGys 312
971 TTTCCGTTATCGACAGAGCGCGCAGCAAGAGCTGTTGCGCTGC..... 1014
   |||::: |||::: |||::: |||::: |||::: |||:::
312 erSerleuAlaGluaspasnGlnSerSerTyrserrarglypHeaspMet 328
1014 ..... 1014
329 thrTythrGlnpHeaspTyrsleucysasnGluvalasPvalth 345
1015 .GTTGCGCGCAGCGCGACCAATATCTCCATCAGCGGTACGACCTCGGCC 1063
   ::::: |||::: |||::: |||::: |||::: |||:::
345 rGysSerProLysProAspAlaPheasnProGysGluaspIleMetGlyT 362
1064 ATTTCTGAAAAACAACCTCTTCAAGTTACAGACA.....GCCGTCAAC 1107
   ::::: |||::: |||::: |||::: |||::: |||:::
362 yrAsnIleuLeuArgylleuIletrpPheIleSerIleuAlaIlethr 378
1108 GGTGCGGACCGCGCATGTGCGCATTTGCTACT.....TACGACGCGCT 1151
   |||::: |||::: |||::: |||::: |||::: |||:::
379 GlyAsnIleIleValleuValIleuIlethrSerGlnTyrlLysleuth 395
1152 AATGCCGTAGACATCCCTCCTACCCCTGTTTGGCGCATTAATCGTGC 1201
   ::::: |||::: |||::: |||::: |||::: |||:::
395 rAlpProArgpHeuMetCysasnleuAlaPheAlaaspLeucysIleG 412
1202 GC 1203
412 ly 412
seq_name: /cgn2_6/ptodata/1/aa/6B_COMB.pep:US-09-100-193-2
seq_documentation_block:
; Sequence 2, Application US/09100193
; Patent No. 6153729
; GENERAL INFORMATION:
; APPLICANT: GARY S. STEIN ET AL.
; TITLE OF INVENTION: NUCLEAR MATRIX TARGETING PEPTIDES AND USES THEREFORE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100.193
; FILING DATE:
; CLASSIFICATION:

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/050,104
FILING DATE: 20-JUNE-1997
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: UMM-024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: Internal
US-09-100-193-2

```

```

alignment_scores:
Quality: 84.50 Length: 371
Ratio: 0.512 Gaps: 18
Percent Similarity: 44.474 Percent Identity: 19.137

```

```

alignment_block:
US-09-303-518D-127 x US-09-100-193-2

```

```

Align seg 1/1 to: US-09-100-193-2 from: 1 to: 415

```

```

28 CTGCCCATCGCGGCGAGACCGAGCAAGTC.....ATTATACGCGGC 71
   |||::: |||::: |||::: |||::: |||::: |||:::
89 LeuProvalAlaPheLysValIleAlaLeuGlyAspValProaspGlyTh 105
72 CGCTATACCAAGTCGCGTTCGTTGCGCAAGATATCGGTATGCGGC 121
   |||::: |||::: |||::: |||::: |||::: |||:::
105 rValValThrValMetAlaGlyAsnaspGluAsnTySerAla..... 119
122 CCTGATGAAAGTCAGAGAGCGCGATCCGTCACAAAAGCCCAAGTCTG 171
   ::::: |||::: |||::: |||::: |||::: |||:::
120 .....GluLeuArgAsnAlaSerAlaValMetLysasnGlnValAla 133
172 .TTTGAAGCAAAAAGMATCCGCGC..... 195
   |||::: |||::: |||::: |||::: |||::: |||:::
134 ProPheAsnAspLeuArgPheValGlyArgSerGlyArgLysSerph 150
196 .....GTGGTTCACCGCGCGCGCGTTCAGGCAAAATCCCGC 232
   |||::: |||::: |||::: |||::: |||::: |||:::
150 eThrLeuThrIleThrValPheThrAsnProThr.....GlnValAlar 165
233 CCATCATCGCGCGCAAAAGCGCGTACTTCAGTCGGTGTGATGCGGTT 282
   ::::: |||::: |||::: |||::: |||::: |||:::
165 hrTyHisArg.....AlaIleLysValThrVal 174
283 GAAGGCAACGAGCAATGCACTGCAACGCTACGCG..... 318
   ::::: |||::: |||::: |||::: |||::: |||:::
175 AspGlyProArgGluProArgHisArgGlnLysLeuGluaspGlnTh 191
319 .....CCCGAAGCGTGGCAAACTTAAGCGCGGCGGANGAANTMNGN 358
   |||::: |||::: |||::: |||::: |||::: |||:::
191 rLysProPheProAspArgPheGlyAspLeuGlu..... 202
359 GCAATGTGATCCATCCGTTTGTGACTGGCGCTGACGCTGACGCTTC 408
   |||::: |||::: |||::: |||::: |||::: |||:::
203 .....ArgLeuArgMetArgValThr 209
409 AGCAAAATCCCTGCC.....GTGGA 428
   ::::: |||::: |||::: |||::: |||::: |||:::
210 ProSerThrProSerProArgGlySerLeuSerThrThrHisPheSe 226
429 TGGCGAGCGGTCCGCACTGTCGATGATGCGATGACCAATCGCGCTNG 478
   ::::: |||::: |||::: |||::: |||::: |||:::
226 rSerGlnProGlnThrProIleGlnGlyThrSerGluLeuAsnProPhe. 242

```

```

479 CGCGACAGCCCTGTGGTTGATCAAGAAGCCGCGCATTTTCAGACA 528
    :|||
243 ..SerAspPro.....ArgGlnPheValArg 250
529 NCMNNGCTGATTTAGCCGTTGACCGCGTAATC..... 567
    :|||
251 SerPheSerThrLeuSerThrLeuGlnSerArgPheProAspProAr 267
568 .....CATGCTGTAGG 580
267 glleryltyrthrlyalameSerAlaAlaPheProtyrSerAlaThr 284
581 CAGCTGGCGACAGCGTCCGCTGAAATGCTGCACATCGAACA... 627
    :|||
284 roSerGlyThrSerLeuSerLeuSerValAlaGlyIleSerAlaThr 300
628 .....CATGAATTC..... 636
301 SerArgPheHisThrTyrLeuProProtyrProGlyAlaProG1 317
637 .....GGCGGCCCGCATCCGCGGTTGAGTGCACGACATTCATT 679
    :|||
317 nasGlnSerGlyProPheGlnAlaAsnProSerProtyrHisLeuTyr 334
680 TCATTGAGCCGCTGGTGCAAAACAAACCGTTGGACCATTCATTCAA 729
    :|||
334 YrGlyThrSerSerGly.....SerTyrGln 342
730 GATGTAATTCGATCGACGCTTTGTTGCAACAGCCGCTGCAACCCGA 779
    :|||
343 PheSerMetValAlaGly..... 348
780 GCGGCTGATTCCTTGGGTGCTCTCAATGCAACAAACGCGCTTTCG 829
    :|||
349 .....SerSerSerGlyAspArgSerProThrArgMetLeuA 362
830 GTACCGTTTGGTGCGAAGTATCGCAATTCGCGGCGCAATTCGTT 879
    :|||
362 laSerCysThrSerSerAlaIleSer...ValAlaAlaGlyAsnLeuMet 377
880 GAGCGACAGACACCGCGCTGATTCGCGTATTCGACGCGCGCATTC 929
    :|||
378 AspArgProSer.....LeuGlyGlyGlnSerAspGlyValGluAl 390
930 ACAAGGCGCGCAC 942
390 aaSpGlySerHis 394

seq_name: /cgn2_6/plodata/1/1aa/PCTUS_COMB.pep:PCT-US96-10602-8
seq documentation block:
: Sequence 8, Application PCT/US9610602
: GENERAL INFORMATION:
: APPLICANT: The General Hospital Corporation
: TITLE OF INVENTION: INHIBITION OF HEPATITIS B REPLICATION
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/10602
: FILING DATE:
: CLASSIFICATION:

```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/017,814
: FILING DATE: 20-JUN-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Clark, Paul T.
: REGISTRATION NUMBER: 30,162
: REFERENCE/DOCKET NUMBER: 00786/282001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO.: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 289 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US96-10602-8

```

```

alignment_scores:
  Quality: 84.00      Length: 166
  Ratio: 1.105        Gaps: 13
  Percent Similarity: 45.783  Percent Identity: 29.518

```

```

alignment_block:
US-09-303-518d-127 x PCT-US96-10602-8 ..

```

```

Align seg 1/1 to: PCT-US96-10602-8 from: 1 to: 289

```

```

788 TTGCTTTGGGTGTTCTCAAGTCACAAACAGCGCTTTCGTAACGCT 837
    |||
100 LeuLeuThrPhe.....HisIleSerCysLeu,Thrp 110
838 TTGGTGGCAAAAGTATTCGCAATTCGCGGCGCAATTCGT...TGACGC 884
    |||
110 heGlyArgGluThrValIleGluTyrIleuValSerPheGlyValTrpIle 126
885 AGACACCGCGGATTCGCGTTCGCTGCTTGAACGCGCGATTCACACAAG 934
    |||
127 ArgThrProProAlaTyrArg.....ProPr 135
935 GCGCGACAGATTAATTTGGAGCGTACACATTCAGATTCGTTATGAA 984
    :|||
135 oAsnAlaProIleLeuSerThrLeuProGluThrValValArgArgA 152
985 GAAGCGCGACAAAGAGCTGCTGGCTGGCGCGACCGGACAA 1034
    |||
152 rg.GlyArgSer.Pro.....ArgArgArgThrP 161
1035 ATATCCATCATCAGCGGTACGACCCCTCGGCCATTTCTTCAAAAACAACTCT 1084
    :|||
161 roSerProAlaArgArgArg.....SerGlnSer 170
1085 TCAAGTTCACGACACCGCTCAGCGTGCGACCGCGCATGGTGGCGAT 1134
    :|||
171 ProArgArgArgArgSerSerThrThrSerThrGlyPro...CysArgTh 186
1135 GGTACTTACGAGCGCGTAATCCGCTAGACATCTGCTGCTGCTTT 1184
    :|||
186 rCysMetThrThrAla.....GlnGlyThrSerMetTyrProSerCysC 201
1185 GCGCGATTTAATCGTCG.....GCGATACGACAA 1213
    :|||
201 yScyThrIlyProSerAspGlyAsnCysThrCysIleProIleProSer 217
1214 GCGCGCAAGCATTTGGGT.....GCTTGAATTCG 1243
    :|||
218 Ser.....TrpAlaPheGlyLysPheLeuTyrPgluTrp 228

```

```

seq_name: /cgn2_6/plodata/1/1aa/5B_COMB.pep:US-08-467-822-21

```

```

seq_documentation_block:
; Sequence 21, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thibierge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flunegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495, 0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4400
; TELEFAX: (202) 408-4000
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 569 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..569
; OTHER INFORMATION: /note="URE B - FIGURE 3."
;
US-08-467-822-21

alignment_scores:
Quality: 83.50 Length: 258
Ratio: 0.898 Gaps: 10
Percent Similarity: 36.047 Percent Identity: 21.318

alignment_block:
US-09-303-518D-127 x US-08-467-822-21 ..
Align seg 1/1 to: US-08-467-822-21 from: 1 to: 569

568 CATGCTGTAAGGAGCTGGCGAGACGCTG..... 597
|||||
333 HistLeuAspLysSerIleLysGluAspValAlaInPheAlaAspSerArgIle 339

```

```

598 ....CGGTGTGAATGCTGCACAATCGAACATCGAACATGATTCGGCGGCC 643
|||||
339 eaArgProGlnThrIleAlaIleGluAspGlnLeuHisAspMetGly.... 354
644 CGCATCCGGCCGGTTGATGGGACGACGACATTCATTTCAATTGACCGCGTC 693
354 ..... 354
694 GGTCAACAACAACCCGTTTGACCACATTCATTCAGATGATTCGAT 743
|||||
355 .....IlePheSerIleThrSerIleSerIleAlaIle 366
744 CGAGCGTTGTTGTCACACAGCGCGTTCGAACACCGACCGCGTATGCTT 793
|||||
366 tGlyArgVal.....GlyGluValIleThr..... 374
794 TGGGTGGTTCTGACATCAACAAACACCGCTTCGCTACCGTTTGCT 843
|||||
375 .....ArgIleThrProGlnThr 379
844 CGGAAGATATCGCAATTTACTCGCGGCAATTCGTT.....GA 881
|||||
380 AlaAspLysAsnLysLysGluPheGlyArgLeuLysGluLysGlyAs 396
882 CGCAGACCAAC.....CGCGTATTCGCGTTCGCTATTCGACG 919
|||||
396 PAsnAspAsnIlePheArgIleLysArgIleSerIleThrIleAsn 413
920 GCGCGATTACACAGCGCGGACGATTTTGGAGCGGTACACAAATCAG 969
|||||
413 tGlyIleAlaHisIleSerAspTyrValGlySer..... 425
970 ATTTCGTTATCGAAGAGCGCGACGAAAGCGTTCGGCTGG..... 1014
|||||
426 .....ValGluValGlyLysTyrAlaAspLeuValLeuTyrSerPr 439
1014 ..... 1014
439 AlaIlePheGlyIleLysProAsnMetIleIleLysGlyLysPheIle 456
1015 .....GTGGCGCGGACGCG 1029
456 ILeuSerGlnMetLysAspAlaAsnIleSerIleProThrProGlnPro 472
1030 GACAATATCTCCATCAGCGGTCAGACCCCTCGGCTTCGTGAACAA 1079
|||||
473 ValTyrTyr.....ArgIleMetPheGlyHisIleGlyAsnLys 486
1080 ACTCTTCAAGTTCACGACACCGCTCAAC..... 1107
|||||
486 sPheAspThrAsnIleThrPheValSerGlnAlaIleAlaTyrLysAlaGly 503
1108 .....GTTGGCGACCGCGCATGTCGCGATTCGATTCATTC 1143
503 IeLysGluGluLeuGlyLeuAspArgValAlaIleProValLysAsnGly 519
1144 GAGCGCGTAAATGCCGCTAGACATC 1167
|||||
520 ArgAsnIleThrLysLysAspLeu 527

seq_name: /cgn2_6/plotdata/1/1aa/6B_COMB.pep:US-08-432-697-21

seq_documentation_block:
; Sequence 21, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thibierge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

```



```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495, 0137-02000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Protein
LOCATION: 1..569
OTHER INFORMATION: /note="URE B - FIGURE 3."
US-08-466-248-21

```

```

alignment_scores:
Quality: 83.50      Length: 258
Ratio: 0.898       Gaps: 10
Percent Similarity: 36.047   Percent Identity: 21.318

```

alignment_block:
US-09-303-518D-127 x US-08-466-248-21 ..

Align seg 1/1 to: US-08-466-248-21 from: 1 to: 569

```

568 CATGTGTAAAGCAGCTGGCCGACGCTG..... 597
      |||||  |||||
333 H1sleuAaplyserIlelysglnAapAlnphelaaspsersArg11 339
598 ....CCGCTGAAATGCTGCCAATGCAACATGATTCGGCGCC 643
      |||||  |||||  |||||  |||||
339 eArgProGlnhrIleAlaAlaGlnaspGlnleuHisaspwetyl.... 354
644 CGCATCCGCGCGTTGAGTGCGACGACATTCATTGAGCCGCTC 693
354 ..... 354
694 GGTGCAAAACACCGTTTGACCATTCATCAAGTGAATGGCAT 743
      |||||  |||||  |||||  |||||
355 .....IlepheserIlehrSerSeraspSerGlnAlaIle 366
744 CGAGCTTTGTTGCAACAGCGCTGTGAACACGCGCGTGAATGCTT 793
      |||||  |||||  |||||  |||||
366 tGlyArgVal.....GlyGlnValIleThr..... 374
794 TGGGTGTTCTCAAGTCAACAAACACGCTTTCGACCGTTTGGGT 843
      |||||
375 .....ArgThrTrpGlnThr 379
844 GCGAAGTATCGCAATACCTCGGCGCAATGTT.....GA 881
      |||||  |||||  |||||  |||||
380 AlaasplysAsnLysLysGlnpHeGlyArgLeuLysGlnLysGlyAs 396
882 CCGACAGAAC.....CCGCTGATTTCCGTTCCGATTCGACG 919
      |||||  |||||  |||||  |||||
396 pAsnspAsnPhenArgLysArgTyrIleSerLysTyrThrIleAsn 413
920 GCGGATTAACACAGCGCGACGATTAATTGGAGCGGTACCAATCAG 969
      |||||  |||||  |||||  |||||

```

```

413 roGlyIleAlaHisGlyIleSerAspTyrValGlySer..... 425
970 ATTCCGTTATGAGAGAGCGCGACCAAGAGCTGTTCGCTGG..... 1014
      |||||  |||||  |||||  |||||
426 .....ValGlnValGlyLysTyrAlaAspLeuValLeuTrpSerPr 439
1014 ..... 1014
439 oAlaPhenpHeGlyIleLysProAsnMetIleIleLysGlyGlyPheIleA 456
1015 .....GTTCGCCGCGACCGC 1029
456 lAeuSerGlnMetGlyAspAlaAsnAlaSerIleProThrProGlnPro 472
1030 GACAAATACCTCCATCAGCGGTACGACCCCGGCATTCCTGAAAAACA 1079
      |||||  |||||  |||||  |||||
473 ValTyrTyr.....ArgGlnMetPheGlyHisHisGlyLysAsnLys 486
1080 ACTCTCAAGTTCACGACGACCGCTCAAC..... 1107
      |||||  |||||
486 sPheAspThrAsnIleThrPheValSerGlnAlaAlaTyrLysAlaGlyI 503
1108 .....GGTGGCGACCGCGCATGGTCCGATGGTACTTAC 1143
      |||||  |||||  |||||  |||||
503 lElysglnGlnLeuGlyLeuAspArgAlaAlaProProValLysAsnCy 519
1144 GAGCGCGTATGCGCGTACGATC 1167
      |||||  |||||
520 ArgAsnIleThrLysLysAspLeu 527

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-09-199-637A-351

seq_documentation_block:

```

Sequence 351, Application US/09199637A
Patent No. 6353411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-Wah
APPLICANT: Cao, Hui
APPLICANT: Drenkard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
FILE REFERENCE: 00786/361002
CURRENT APPLICATION NUMBER: US/09/199,637A
PRIOR APPLICATION NUMBER: 60/066,517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 351
TYPE: PRF
LENGTH: 447
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-351

```

```

alignment_scores:
Quality: 83.00      Length: 277
Ratio: 0.741       Gaps: 14
Percent Similarity: 40.433   Percent Identity: 25.271

```

alignment_block:

US-09-303-518D-127 x US-09-199-637A-351 ..

Align seg 1/1 to: US-09-199-637A-351 from: 1 to: 447

```

15 AAAAGCTTAACCTGCCCATCGCGGCGACAGC.....GGACAGAGTCA 58
      |||||  |||||  |||||  |||||
210 ArgArgProAlaAspGlyHisAlaGlyGlnThrAlaProGlyAspProG 226
      |||||  |||||  |||||  |||||

```

```

59 TTTATGACGGCCCGTCATTTACCGAAGTCGGTGGCGGAGAAATAT 108
    ||||| ||| :||| |||||
226 yysglaralargllyargarproalagllyargararglthrc 243
    ||||| ||| :||| |||||
109 GC.....CGTATGCCCGCCCTGATGAAAGTCAA 137
    ||||| ||| :||| |||||
243 yalaserSerAlaProProthrarThrSerProPolysSerAlaPro 259
    ||||| ||| :||| |||||
138 GGAAGGCGATGCCGTCAAAAAGGCCAAGTGCTTTGAGACAAAAGN 187
    ||||| ||| :||| |||||
260 GLyAlaserAlaThrSerThrThrAlaser.....Th 271
    ||||| ||| :||| |||||
188 ATCCGGCGCTGCTTTACCGCCGCTTTACAGCAAAATCGCCGATC 237
    ||||| ||| :||| |||||
271 rSerSerSerCysAlaThrThrAlaAla...ArgThrProArgLysHsp 287
    ||||| ||| :||| |||||
238 CATCGCGGGAAGCGGCTACTTCAGC...GTCGTGATTCGGCTTGA 284
    ||||| ||| :||| |||||
287 roAlaAlaThrThrThrHisProGlnAlaProGlyArg..... 299
    ||||| ||| :||| |||||
285 AGCAACGACGAATCGATTGAGTTCAGCGCCGCGCGGCTTGCGCA 334
    ||||| ||| :||| |||||
300 .....ArgHisArgProAlaGlyArg 306
    ||||| ||| :||| |||||
335 ACTT.....AAGCGCGANGAANTNNGCAATCTGATCCAAATCCGCT 378
    ||||| ||| :||| |||||
306 gGlnAlaAspArgArgArgThngLylAlaGlnLylLeuProLeu... 321
    ||||| ||| :||| |||||
379 TTGTGACTGGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 428
    ||||| ||| :||| |||||
322 .....ProGlyGlnArg 325
    ||||| ||| :||| |||||
429 TGCCGAGCGTTCGCCATCTT.....CG 451
    ||||| ||| :||| |||||
326 ProArgAlaGlnLylHisAlaGlnAlaArgLeuThrProValaArgArg 342
    ||||| ||| :||| |||||
452 TCAATGCGATGGACACCAATCCGCTNGCGGAGACCCCTGTTGTGATC 501
    ||||| ||| :||| |||||
342 gProAspProAlaSerArgProAlaProGlyArgCysAlaGlnLysGlnP 359
    ||||| ||| :||| |||||
502 AAAGAGCGCGNCA..... 515
    ||||| ||| :||| |||||
359 roGlyArgArgArgGlnProGlnArgAsnArgGlnProArgGlyLeuPro 375
    ||||| ||| :||| |||||
516 .....NGATTTCAGACGANGTNGCTGCT.....AT 541
    ||||| ||| :||| |||||
376 GLyArgHisArgAlaGlnAlaAspHisAlaGlnLylThrArgGlyAspProLe 392
    ||||| ||| :||| |||||
542 TGAGCCGTTTGACCGAGCGTAAATCCATGTGTGAAGGCGAGCTGGCGCA 591
    ||||| ||| :||| |||||
392 uGlnProHisArgArgGly.Pro.....AlaProGlyPro 403
    ||||| ||| :||| |||||
592 GAGGTGCGCGTCTGAAATGCTGCCAACAATGAAACACATGATTCGGCGG 641
    ||||| ||| :||| |||||
404 AspValPro.LeuAspAlaLeuProProGlnLylHisArgLeuL 420
    ||||| ||| :||| |||||
642 CCCGATCCGCGCGCTTGAGTGGACGACGACATCATTTATTCAGCGCG 691
    ||||| ||| :||| |||||
420 ySValLysArgProValaArgArgGlnAlaThrThrPheSerLeuLeuArg 436
    ||||| ||| :||| |||||
692 TCGGTGCAACAAACCGTTTGACCA 718
    ||||| ||| :||| |||||
437 GLyAspGlnProGlyArgArgGlyPro 445
    ||||| ||| :||| |||||
seq_name: /cgn2_6/prodata/1/1aa/6B_COMB.pep:us-09-105-537-2
seq_documentation_block:
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.

```

```

; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methylein and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-2

alignment_scores:
    Quality: 83.00      Length: 334
    Ratio: 0.529      Gaps: 16
    Percent Similarity: 47.006      Percent Identity: 21.856

alignment_block:
US-09-303-518d-127 x US-09-105-537-2 ..

Align seg 1/1 to: US-09-105-537-2 from: 1 to: 5215

118 CGCCCTNGATGAAGTCAAGAGGAGATGCCGTC..... 153
    ||| :||| ||||| |||||
1695 ArgLeuMetClnSerAlaArgGlnGlyAlaMetIleAlaValGlnAl 1711
    ||||| ||| :||| |||||
154 .AAAAAGGCGCAAGCTGTTTGAGACAAAAGNATCCGCGCTGCTGT 202
    ||||| ||| :||| |||||
1711 agLylGlnAlaGlnValAlaGlnSerLeuLysGlyTyrGlnGlyArgVala 1728
    ||||| ||| :||| |||||
203 TTACCGCGCGCTTCAGGC..... 222
    ||||| ||| :||| |||||
1728 laValAlaValaValaValaValaValaValaValaValaValaVala 1744
    ||||| ||| :||| |||||
223 .....AAATCCCGCATTCATCCGCGGAGAAAGCGC..... 255
    ||||| ||| :||| |||||
1745 AspAlaAlaGlnLylLeuArgAlaValaValaValaValaValaVala 1761
    ||||| ||| :||| |||||
256 .....G 256
    ||||| ||| :||| |||||
1761 rArgArgLeuArgValaSerHisAlaPheHisSerProHisMetAspArg 1778
    ||||| ||| :||| |||||
257 TACTTACGTCGCTGCTGATGCCGTTGAGGAGCAAGCAATTCAGTTC 306
    ||||| ||| :||| |||||
1778 alLeuAspGlnPheLeuArgValaAlaGlnLyl.....LeuThrPhe 1791
    ||||| ||| :||| |||||
307 GAAGCTACGCGCCGCGAAGCTTGCAAACTTAAGCGGCGANGANTNG 356
    ||||| ||| :||| |||||
1792 GlnGlnProArgIleProValaValaSerThrValaThrGlyAla..... 1805
    ||||| ||| :||| |||||
357 NNGCAATCGATCAATCCGT.....TTGTGAGTGTG 388
    ||||| ||| :||| |||||
1806 .....LeuValaThrSerGlyLylLeuThrSerProAlaTyrTrpVala 1820
    ||||| ||| :||| |||||
389 CGCTGCTGATGCTGCTGACCAAAATCCCTGCGGTGATGCCGAGCG 438
    ||||| ||| :||| |||||
1820 spGlnIleArgArgProValaArg..... 1827
    ||||| ||| :||| |||||
439 TTGCGCATTCGTCATGCGATGAGACCAATCCGCTNGCGGAGACCC 488
    ||||| ||| :||| |||||
1828 .....PheLeuAspAlaValaArgThr.....LeuAlaAlaGlnAs 1839
    ||||| ||| :||| |||||
489 TGTGCTGTGATCAAGAAGCCGCGANGATTCAGACGANGTNGCTGG 538
    ||||| ||| :||| |||||
1839 palThrValaLeuValaGlnIleGlyProAsp.....AlaVal 1851
    ||||| ||| :||| |||||
539 TATGAGCGCTTTGACGAGCGTAAATCCATGTGTGAAGCGACGCTGC 588
    ||||| ||| :||| |||||
1851 alLeuThrAlaLeuAlaGlnGlnAlaLeu.....AlaProGly 1863
    ||||| ||| :||| |||||

```



```
589 GCAGAGTCGCCG...TCTGAAATGCTGCCAATCATGCAATGCAAT 635
    ::|||::||| ::::: :::::
1864 ThrAspAlaProAspAlaArgAspValThrValValProLeuLeuArgAl 1880
    ::|||::||| ::::: :::::
636 CGCGGCGCGCGCATCCG.....CCCGTTTGAGTGGCACGCACACA 673
    ||| |||::||| ::::: :::::
1880 aglyArgProGluProGluThrLeuAlaAlaGlyLeuAlaThrAlaHisV 1897
    ::|||
674 TTCAT.....TTCATTGAGCGCGTCGTCGCAAC 702
    ::||| ::||| ::|||
1897 aHisGlyAlaProLeuAspArgAlaSerPhePheProAspGlyArgArg 1913
    ::|||
703 AAACCGTTTGAGCATCATCAATTATCAAGATGTAAATGCCATCGACGTTT 752
    ::||| ::||| :::::
1914 ThrAspLeuProThrTyrAlaPheArg.....GluHisTyr 1926
    ::||| ::||| :::::
753 GTTGGCAACAGCGCGCTGTACACCGGCGCGTGTGCTTGGGTGGT 802
    ::||| ::||| ::||| ::|||
1926 rTyrLeuThrProGluAlaArgThrAsp...AlaArgAlaLeuGlyPheA 1942
    ::|||
803 CTCAGTCAACAACCAACCGCTCTTGGCGTACCGTT...TTGGGTGCGAAA 849
    ::||| ::||| ::||| ::|||
1942 spProAlaArgHisProLeuLeuThrThrValGluValAlaGlyGly 1958
    ::||| ::||| ::||| ::|||
850 GTATCGCAATTAATCTCGGCGCGAATTGTTGACGACACAAACCGGTGAT 899
    ::||| ::||| ::||| ::|||
1959 AspGlyValLeuLeuThrGlyArgLeuSerLeuThrAspGlnProTyrPle 1975
    ::|||
900 TTCGCGTTCGATATGTAACGGCGCATTTACACAAAGCGCGCAGATTATT 949
    ::||| ::||| ::||| ::|||
1975 uAlaAspHisMetValAsnGlyAlaValLeuLeuProAlaThrAlaPheL 1992
    ::|||
950 TG 951
    ||
1992 eu 1992
```

